CYTOCHROMES C AND THE EVOLUTION OF GRAM POSITIVE BACTERIA.

A thesis presented to the University of Edinburgh
for the degree of Doctor of Philosophy
by
Guðmundur Óli Hreggviðsson

March 1993

Head of Department: Professor Richard P. Ambler
Supervisor: Richard P. Ambler.

Institute of Cell and Molecular Biology
University of Edinburgh

I hereby declare that this thesis has been composed by myself and that except where stated, the work described is my own.

ABSTRACT.

I. Purification, characterisation and structural analyses of cytochromes c from *Bacillus azotoformans*.

The content and cellular location of cytochromes in *Bacillus azotoformans* was examined under aerobic and denitrifying anaerobic conditions. Two cytochromes were induced under aerobic conditions, one of which was only expressed under conditions of high aeration. One cytochrome, a b-type cytochrome, may be induced under denitrifying conditions. Cytochromes common to both pathways appear to be expressed in similar amounts per unit cell weight. Both c- and b-type cytochromes from *B. azotoformans* are membrane bound. The haem domains of four different c-type cytochromes, designated P1-c552, P2-c551, P3&P4-c555 and P5-c552, were isolated by mild proteolytic treatment and purified. Properties and other data on these cytochromes are given in the table below. Sequence information was obtained on all four cytochromes. Two of them may be of different structural type than previously sequenced from the genus. The least information was obtained on the cytochromes P2-c551, and P5-c552, respectively, 10 and 37 amino acids from the N-terminus. Peptides were sequenced, which appear to cover the greater part of the cytochrome P3&P4-c555. The haem domain polypeptide of P1-c552 was completely sequenced by manual and automated techniques of proteolytic peptides. Similarities and differences of the three dimensional fold of the cytochrome P1-c552, as deduced from primary sequence characteristics, to *Pseudomonas aeruginosa* cytochrome c 551 (a subclass ID cytochrome c), of known crystallographic structure, are evaluated. It is proposed that the lesser shielding of the haem in the P1-c552 cytochrome and other homologous *Bacillus* cytochromes c may explain their lower midpoint redox potentials. An overall alignment of Class I cytochromes c is presented which differs from previous ones in many aspects. Revisions are suggested to the existing classification schemes of cytochromes c of class I and taxonomic ordering of cytochromes c is proposed at a higher taxonomic level than currently is done. It is argued that the cytochromes

P3&P4-c555 and P1-c552 have closest relatives outside the phylum, respectively, in the c4 (subclass IC) cytochromes and the ID subclass cytochromes c found in the β– and γ-subdivisions of purple bacteria. It is also proposed that a cytochrome c, fused to cytochrome c oxidase in the thermophilic strain *Bacillus* PS3 (Ishizuka *et al.*, 1990) forms a homologue group with c2 sequences found in the α-subdivision of purple bacteria. An amino acid signature of cytochrome c is deduced from an overall alignment. The cytochromes of the different taxonomic groups adhere to this pattern,to a varying degree. The evolutionary implications of the distribution of the proposed cytochrome c groups (taxons) in the phylogenetic eubacterial-16S rRNA tree, are discussed. The electron transport pathway in the genus *Bacillus* is compared with its counterparts in distantly related bacteria. The significance of observed similarities and differences in types of cytochromes and composition for the elucidation of the evolutionary history of the Gram positive phylum as well as for eubacteria in general, is emphasised. Cloning of cytochromes c552 from *Bacillus licheniformis* was attempted but proved unsuccessful. Different probing methods were compared. An approach using a single long probe based on most probable codon usage and washing of hybridisation filters with methyl ammonium chloride proved successful for detecting the gene. The gene appears to reside on a 2.7 kD fragment of Hind III restricted genomic DNA:

Data on cytochromes c obtained from *B. azotoformans*.

	Cytochromes				
	P1	P2	P3	P4	P5
Mol. weight (kD) (a. acid composition)	8,9	9,3	11,3	10,3	13,7
(SDS-PAGE)	8,9	<8	<8	<<8	>9 <12
N-terminal amino acid	Gly	Gln/Glu	Gln/Glu	Gln/Glu	Gln/Glu
Absorption maxima (nm, reduced): α	552	551	555	555	552
β	523	522	523	523	524
(Soret) γ	411	409	417	417	412
Extinction ratios 280 nm/α_{red}	0.70	0.84	N.D	0.72	0.78
α_{red}/β_{red}	1.61	1.69	N.D	1.38	1.66
$Soret_{red}/\alpha_{red}$	5.61	6.21	N.D	6.56	5.14
Redox potential	122	93	N.D	124	188
Conditions of expression	Aer*/An*	Aer/A	Aer/An	Aer/An	Aer/An
Cont. sequence** (number of a. acids)	82	10	***24,	24, 28	37
Class	ClassI	ClassI	ClassI	ClassI	ClassI
Proposed: related subclass (or type)	ID	ID??	IC?	IC	(c2)

* Aerobic/Anaerobic growth conditions. ** Continuos sequence as deduced from overlapping peptides and amino acid composition. *** P21 and P4 were pooled, three continuos sequence areas of different length were obtained with no overlap between them.

TIL BRYNJU

'Perhaps all that is needed, to put Humpty Dumpty together again, is more horses and more men'.

Richard P Ambler, 1991

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENT.

I wish to thank my wife Brynja Ingadóttir for her constant encouragement and support. I am grateful to Professor Richard P. Ambler for his supervision of the research, for providing facilities and materials for the experimental work and for his valuable criticisms of the thesis, the patience, understanding of, and help in matters that delayed the presentation of this thesis. Very special thanks to Margaret Daniel for her friendship, her patient tutoring in the techniques of protein sequencing and all other assistance in the laboratory. I am grateful to Professor Martin D. Kaman for the invitation to the inspiring conference on bacterial cytochromes c at the National Institute of Health in Maryland, USA. I thank Dr Graham Pettigrew for the help in measuring the redox potentials and valuable comments, and the people of the Welmet Protein Characterisation Facility, especially Steven Peacock for introducing me to automated sequencing. I thank Dr. Patrik Holt in Heriot Watt for his considerable help in taming the computers and stimulating 'symposiums'. I thank the many people in the department, Dr. D.Finnegan, Dr. N.Murray, Dr. P. Ford, Dr A. Coulson and Dr. J. Collins to name but few, for their advise and helpful discussions and for supplying enzymes and other materials. I thank my present colleagues and especially Dr Jakob K. Kristjánsson for their encouragement during the period of writing the thesis. I thank Patrik and Sigga Holt, Emil and Didda, Daniella and Sid, Gill and Ian, Reimar and Hanna Gréta, Bjössi and Kristveig, Tom, Katrin, the very SCOTTISH people in Candlemakers´ Arms and all the others for making my stay in Edinburgh (above and beneath the bridges) so enjoyable. I thank my parents for their encouragement and for looking after my two daughters whenever needed in the evenings and during weekends the past year. My two daughters, which were born during this period gave me a whole new perspective on all things. I thank them.

CHAPTER 1
1. INTRODUCTION

1.1. BACKGROUND

Information, drawn from primary sequences of macromolecules, has evoked speculation on many topics (Anfinsen, 1959). It has opened up new and vast fields of study, such as the study of structural/functional relationships of proteins and has had a great impact on other, already well established fields, such as the study of phylogenetic relationships of living organisms.

Primary structures, in conjunction with other information, are important for defining the characteristics that determine the biochemical properties, and functional mechanisms of proteins. These latter in turn, prescribe the constraints imposed on the evolutionary course of the protein structure. The extensive sequencing and crystallographic studies on cytochromes c have, in this context, been very valuable in forming current ideas on protein structures and their evolution (Fitch and Margoliash, 1967; Salemme, 1977; Dickerson, 1980; Chothia and Lesk, 1984, 1986, 1987; Moore and Pettigrew, 1990).

On the basis of the early extensive cytochrome c sequencing work from eukaryotes, a theoretical foundation was laid for phylogenetic studies at the molecular level. Since then, the sequencing and analysis of the universally occurring molecule, 16S rRNA, from bacteria, has been fundamental in suggesting meaningful phylogenetic relationships amongst the higher taxonomic orders of bacteria, previously more or less indecipherable (Zuckerkandl and Pauling, 1965; Ambler, 1985; Woese, 1987, Woese *et al.* 1987).

By comparing phylogenetic trees, based on different genes or other macromolecular sequences, the concept of bacterial species and ranking of higher taxa has been put in appropriate focus (Woese, 1987; Ambler, 1991; Kandler, 1985; Sonea, 1989).

Sequence comparisons of cytochromes c have also proved to be invaluable in attempts to unravel the evolutionary history of metabolic pathways (Ambler, 1991; Dayhoff *et al.*, 1978; Jones, 1985; Moore and Pettigrew, 1990).

The references cited above give an overview of the current ideas in the field of sequence analysis. In many points they differ, sometimes only in nuances or vantage points, but at other times, conclusions and opinions do conflict. In the following sections of this chapter, an attempt will be made to make an assessment of the current status of the study of macromolecular sequences to present the results, theories and views of the various workers in the field. Present ideas based on sequence analyses, about phylogeny and metabolic substructure of the phylum of Gram positive bacteria are presented.

1.1.1 Neutral versus selective evolution

It was realised early on, that protein sequences might be used for phylogenetic and taxonomic purposes (Zuckerkandl and Pauling, 1965). They proposed, that the rate at which mutations accumulated was more or less constant (a molecular clock), and therefore that it should be possible to deduce an approximate time scale as well as phylogenetic branching order from sequence divergence data. This led to the development of the theory of neutral mutation (Kimura, 1983), which postulated that sequence divergence is a stochastic process and evolution proceeds by an accidental sampling of possible neutral mutations.

When it was discovered that cytochrome c was universally distributed in eukaryotic mitochondria, the possibility of putting this hypothesis to the test became feasible. Sequencing of cytochromes c from a wide range of organisms demonstrated, that the phylogenetic relationships were generally in concordance with the phylogenetic tree, based on palaeontologic evidence (Dickerson, 1971), and appeared thus to support the hypothesis of the molecular clock.

Comparative studies of mitochondrial cytochrome c sequences convinced many people that the majority of mutations were selectively neutral, and that natural selection only acted by weeding out deleterious mutations. Thus, natural selection asserted itself only negatively. The hypothesis was also supported by early experiments of Margoliash *et al.* (1976), who were unable to find functional differences between

mitochondrial cytochromes c, considerably divergent in primary structure. The results of shotgun mutagenesis experiments of cytochrome c in yeast by Hampsey *et al.*; (1986) are also relevant. In these experiments, mutations were selected on the basis of deleterious or impaired function. The mutations almost invariably turned out to be in invariant or conserved positions as deduced by multiple alignment. Residues occupying these positions are considered to play an important part in the folding and interior packaging of secondary structures. In both these sets of experiments, the results are not conclusive and the theory of neutral mutation cannot be said to be proven. The assay used in the former experiments were not carried out under physiological conditions and not sensitive enough to detect subtle kinetic differences. Alterations in assay conditions now, allow kinetic differences to be detected between cytochromes c that only differ from each other in one or two positions (Margoliash *et al.*, 1976). In the latter experiments, the criteria for mutation selection may have been too severe to allow for the isolation of near fully functional, wild type equivalents.

Subtle kinetic differences may assert themselves more selectively in the natural environment of the organism (which is constantly changing for unicellular organisms), than is observed under artificial laboratory conditions. Also, evolution works on many more properties of protein than purely functional ones (optimisation of enzymatic activity, rate/substrate specificity). Thus, if maximum activity of a particular protein, is not the rate limiting step in a biochemical pathway, it will not be the most important factor steering the evolutionary course of the protein (Moore and Pettigrew, 1990).

On the basis of the present, increasingly detailed, analyses of lineages in eukaryotic proteins and more distantly related prokaryotic sequences, the consensus now emerging, is that molecular evolution is a quasi-clock-like process (Ochman and Wilson, 1987). Variable mutation rates are observed both between and within lineages of a macromolecules and even in different regions of the same macromolecule (Woese, 1987). These observations suggest, that evolutionary change is a more complicated process than can be accounted for by the theory of neutral mutation in its simplest form, which consequently implies that natural selection may play a more significant role than previously has been accepted.

Without discarding the theory of neutral mutation, the evolutionary process may be explained as largely proceeding by random fixation of neutral mutations, punctuated by periods of natural selection and

therefore adaptive fixation rates. The concept of the covarion (Fitch and Margoliash, 1967,) can be used to construct a model of the evolutionary process according to this modified neutral theory. The following definition of covarions in mitochondrial cytochromes c, is based on a representation by Moore and Pettigrew (1990). The fraction of the gene that is theoretically invariable in mitochondrial cytochromes c, at particular nodes in a maximum parsimony tree, is found to be 90% in an individual species. The remainder of the residues, 10% or so, are the concomitantly variable codons able to accept change without detrimental effect to the protein and are defined or designated covarions. Previously, it was assumed, that since 75% of cytochrome c residues have been observed to change during eukaryotic evolution, this is the fraction of residues on which neutral selection is able to act on in each individual protein. A measure of the constraints on a particular protein is therefore the size and composition of the covarion group (10 amino acids for the cytochrome c above) and this is reflected in the fixation rate of mutations. Fast evolving protein, such as fibrinogen has a large size covarion group, whereas more constrained proteins, such as mitochondrial cytochrome c has a relatively small number of covarions and, therefore, a comparatively slow evolutionary rate.

The concept of the covarion can be linked to the hypothesis that proteins at any time are slightly suboptimal in their key properties and not able to make use of all advantageous residues at once. When it is optimised in one attribute, it necessarily becomes suboptimal in another (Zuckerkandl, 1976). Change at one covarion can therefore cause both creation of new covarions and removal of old. The size of the covarion group remains approximately the same, while the evolutionary pressure is constant and, therefore, also the rate of change. When function, structure and/or environment changes, the protein may be subject to rapid directed selection and adaptation, during which the covarion group may alter size and composition and may result in a different rate of change afterwards.

McLachlan (1987) has speculated on the role of minor and major mutational events. He hypothesises that small steps, single point mutations and small insertions and deletions of just few amino acids at loops, may lead to changes in substrate specificity or activity. He suggests that this will allow new ligands to bind, or subunits to aggregate, but the overall domain structure will be left almost unaltered. In contrast, major mutational events, massive deletions and insertions, or new splicing or initiation sites in introns and extrons, will open up new evolutionary

avenues. Such changes realise new possibilities, latent in the original component structural elements. Without these major mutational events, proteins would reach an evolutionary stagnation point (optimum evolution?), in which every point mutation is slightly deleterious (McLachlan, 1987). From the vantage point of neutral mutation theory this could be postulated to be the stage where most mutations in a protein are neutral. The specific structural protein type at this stage can be envisaged as a set of sequences, occupying the same tight corner of sequence space.

On the nucleotide level, the theory of neutral mutation in its "purest" form may apply to the silent third base in codons and indeed it appears that the mutation rate is faster at these sites. This indicates that there are no or fewer selective constraints imposed on these sites than on the more determinative ones. However, these sites may not be entirely exempt from natural selection. The third base can be a factor contributing to the control of the rate of expression, and may have a selective value or play some vital integrated part in protein translation, concomitantly with ribosomal rRNA and tRNA. The nature of the latter could be species specific and could be related to the distinct base composition of the organism.

Theories explaining evolutionary change solely by natural selection are at present less well developed. The most recent is the attempt of Gillespie (1986). He argues that randomly varying selection coefficients, rather than neutral mutations, account for molecular evolution with a time scale period of 10^3 years. A statistical treatment is presented, where he shows that the episodic model fits sequence divergence data and can account for variable mutation rates within and between lineages. The observed evolutionary change in the extant species can then be explained by introducing the concept of an episodic clock. Gillespie postulates, that evolution may proceed by bursts of substitutions separated with longer periods without any substitutions. The apparent constancy may therefore be due to averaging out of periodical rapid change.

Finally, the more practical view of Felsenstein, (1988) should be mentioned. He argues that it does not matter whether nucleotide substitutions are neutral or selective. The inability of resolving the controversy is an advantage since the neutral mutation theory can then be used as if it were true and other theories would make indistinguishably different predictions from the data. However, he advocates that methods of evaluating consistency and reliability of trees need further development.

1.1.2. Phylogenetic Reconstruction

The macromolecule that is optimal for phylogenetic studies should meet the following criteria (Pace, 1985 ; Ambler, 1985):

1) It should not be subject to appreciable lateral transfer.
2) It should be universally distributed.
3) The function and structure of the molecule should be the same in different species, ensuring a constant rate of nucleotide/amino acid substitution.
4) The genes involved should provide for a sufficiently slow 'clock', i.e. the change in the sequence must be slow enough, so that the largest of the genealogical distances can be detected.
5) The detection and isolation of the macromolecule concerned, should be feasible.

For eukaryotes, palaeontological evidence, and perhaps, DNA sequences obtained from ancient preserved biological tissue, can be used to assess the plausibility of phylogenetic reconstruction, based on sequence distance data, and also to derive a time scale of the molecular evolution in the lineages. This is not possible for bacteria, and the time scale of bacterial evolution, as deduced from sequence divergence data, at least in principle, is relative. However, independent estimates have been attempted. Ochman and Wilson (1987) have, for example, related the divergence of *Enterobacteriaceae* with the appearance of mammals. Similarly, De Wachter *et al.* (1985) have postulated that stromatolite beds, starting 3.1 x 109 years BP, correspond to the sequence distance between cyanobacteria and other eubacteria.

1.1.2.1. Main methods used in phylogenetic reconstruction

Many different methods of phylogenetic tree construction are now in existence. Each method has its proponents and the controversy centres on the degree of robustness of individual methods in dealing with unequal mutation rates and great phylogenetic distances. The following sections describe those which are most commonly used and are mainly based on reviews by Swofford and Olsen (1990) and Felsenstein (1988).

1.1.2.1.1. Distance matrix trees

Phylogenetic tree construction is based on pairwise sequence differences. Homologous sequences are aligned, the relative importance of each mutation event is estimated and the data is processed to give numerical value for evolutionary distances between the species. The values are then subjected to numerical analysis in order to construct the best phylogenetic tree.

UPGMA

Cluster analysis is a family of methods used to construct phylogenetic trees in this form. The one most commonly used is the UPGMA (pair group method using arithmetic averages, Sneath and Sokal, 1967). The underlying assumption of the UPGMA method when applied to molecular data is that expected mutational rate is constant. Mathematically the distances must satisfy the axiom of the triangle inequality, stating that the distance between **a** and **c** cannot be greater than the sum between **a** and **b** and **a** and **c**. Such distances are called ultrametric. Ultrametric distances emerge in the data set if the mutation rate is uniform in the sequence lineage under study. The resultant tree will be rooted.

The prior assumption that conditions of a molecular clock are fulfilled within and between lineages has been shown to be unsustainable. The method is, nevertheless, conceptually akin to the more complex distance matrix methods. It is fast and in many cases, depending on the distances to be covered, the results are only marginally different from those of more complicated methods (Felsenstein, 1988; Sneath, 1987). In computer simulation studies it has been shown that when distance estimates are subject to large stochastic errors, UPGMA is often superior to other distance matrix methods in recovering the true tree (Nei, 1987)

Additive distances

Trees are constructed, based on the assumption that distances are additive. By using this method the priori prerequisite of a molecular clock, a uniform rate of mutations in the tree, is relaxed (Felsenstein, 1988;). Analysis of this kind aims at finding the geometry and branch lengths, which best fit the phylogenetic distances between the species considered. In the method of Fitch and Margoliash (1967), which is commonly used, the goodness of the fit between the observed and expected distances is estimated by least square method. The aim is to optimise the fit by accommodating different rates along different branches and the best tree

is the one, which minimises the discrepancy between calculated and observed distances. The resultant trees will be unrooted.

The methods are sensitive to unobserved multiple changes, with a possible underestimation of distances between distantly related sequences, and artificially placing fast evolving species together. The reliability of the methods is dependent on the estimation and transformation of distances.

Sneath (1989) emphasises, that sampling error may affect this type of phylogenetic analysis and points out, that the effect is usually overlooked. Thus, assuming that evolution rate is more or less constant and divergent (the conceptual basis of phylogenetic reconstruction methods) it follows that numbers of changes in unit time will have a normal distribution. The standard deviation of changes at an internode can therefore be expected to be close to the root of the differences. Then it follows that the internode may lie within about two standard deviations above and below of the estimated one. As an example, if the estimated internode is 25, the standard deviation is +/- 5 and the internode may be anything from about 15 to 35. The sample error effect has perhaps the most serious consequences for attempts to assign phylogenetic branching order. If many internodes lie within a region of sample error uncertainty, it may be difficult, if not impossible to assign order to the branches with any degree of certainty.

1.1.2.1.2. Maximum Parsimony analysis

This method treats each position individually and attempts to find minimum evolution or maximum parsimony in sequence data. A maximum parsimony tree estimates the relatedness of extent sequences by fewest genetic changes (the shortest tree) and the number of parallel and convergent changes are kept to the minimum.

The method can be broken down into two parts:

1) The calculation of the minimum genetic changes (nucleotide replacements) of any given tree and,

2) The search for the shortest among all possible trees.

The first part has been solved but no simple solution has been found for the second part. In theory, the only way to find the tree with the least genetic change is to compare all the trees possible, from the sequence data. This is not computationally feasible for more than 10 sequences.

Therefore, various shortcuts have been devised, for example, by starting with different plausible trees such as those determined with less computationally intensive methods and proceed from there, by branch swapping, until no tree is found of shorter length after extensive trials. The resultant tree has a high probability of being the shortest tree, even if it can not be proven.

The parsimony methods assume that homoplastic (convergent, reverse and parallel) changes are minimal. However, as the number of changes between sequences increases, the proportion of those being homoplastic increases. These changes cannot be detected in long unbranched lineages, thereby creating the danger of an artificial association of the taxa at the end of these long branches.

1.1.3. Lateral Gene Transfer

The occurrence of lateral transfer is well documented for bacteria (Jones and Sneath, 1970; Krawiec and Riley, 1990), but to what extent it takes place, is still open to speculation. If it has played a significant part, there should be discordance in topologies between phylogenetic trees constructed for different characters, and especially single gene traits. Characters would be found to be erratically distributed among disparate organisms, and several different permutations of homologous characters would be found in a set of related organisms (Sonea, 1989; Ambler, 1985). The consequences of such a mode of evolution in its extreme form, would be that phylogenetic relationships between bacteria would have to be represented by a reticulate network rather than a tree. On the other hand, 'it is possible that the bacterial genome has an inviolate core of genes that code for essential functions, and which are so well matched to each other that no segregation is possible without making them function inadequately. Or more likely, there could be a hierarchy for genes based on the ease which they can be assimilated into a foreign genome' (Ambler, 1985).

Factors that may enable lateral transfer to take place could include the following (Maynard-Smith *et al.*, 1991, Jones and Sneath, 1970; Krawiec and Riley, 1990)

1) The degree of transfer is dependent on the structural integration of the gene product with other molecules. Thus, one could explain the

apparent ease with which the drug resistant factors are exchanged among bacteria by the fact that these are most often single gene encoded characteristics.

2) The structural integration of individual proteins in some complicated multicomponent system does not preclude the possibility that whole clusters of genes that code for some intricate ´machinery´ could be transferred. Such might be the case with the nitrogen fixation (nif) genes based on the following line of reasoning (Ambler, 1985). Nitrogen fixation is widely but erratically distributed among bacteria. It is found in four out of five of the major eubacterial groups. By using nif genes isolated from *Klebsiella pneumoniae* as a probe, homologous sequences were found in 19 widely disparate nitrogen fixating species, but not in any of 10 non-fixating species examined. This fact reflects a very high degree of sequence similarity which is difficult to reconcile with any theory suggesting a common ancestral gene origin. Metabolic genes evolving over such a large time-span would be expected to have diverged beyond recognition by a common probe. This is further supported by the sequence similarity of nitrogen reductase from *Clostridium pasteurianum* and *Azotobacter vinelandii* which turned out to be so great that, to account for it by vertical evolution it would have had to be as slow as that of histone H4. Also, nif genes are known to be plasmid encoded in some species.

One requirement for a multicomponent system to be transferred, is that the genes coding for the various proteins have to be organised in blocks in the donor species. For respiratory proteins, such blocks have been identified in *E. coli* for the subunits of the proton translocating ATPase (Walker et. al., 1984) and for the bc1 complex in *Rhodobacter capsulatus* (Daldal *et al.*, 1986). An indication of such an organisation, may also be found in *Pseudomonas aeruginosa* for the components of the denitrification pathway. A genetic fragment, cloned and sequenced, contained the cytochrome c551 and a part of the cd1 nitrite reductase (Nordling *et al.*, 1990). This latter find may support the hypothesis that lateral transfer has played a part in the distribution of denitrification among bacteria (discussed in Pettigrew and Moore, 1987)

3) Control mechanisms of the foreign genes will have to be compatible with those of the acceptor strain, ensuring their function and/or efficiency. In principle this would decrease the possibility of a successful gene exchange between distantly related species.

4) The acquired trait must be novel and confer a strong selective advantage to the acceptor species. This hypothesis is supported by the rapid spread of antibiotic resistance genes in recent times. It would account for the occurrence of homologous β-lactamases in Gram negative and Gram positive bacteria and for the distribution of the nif genes (Ambler, 1985). It could also explain the apparent similarity of nitrate reductase which plays a part in denitrification, another characteristic which has a wide but erratic distribution among bacteria (Jones, 1985) .

Possible incidences of lateral transfer will be discussed in the appropriate context in the following sections.

1.1.4. Bacterial Phylogeny and Cytochrome C

It is evident that mitochondrial cytochrome c is well suited for phylogenetic studies as it meets, to a considerable extent, the requirements prescribed for an ideal macromolecular chronometer. The situation is somewhat more complicated in the bacterial realm. Bacteria are metabolically more versatile and subject to more direct evolutionary pressures than mitochondria, which are more or less sheltered in eukaryotic organisms. This is probably reflected in the high degree of structural diversity encountered in homologous cytochromes c among bacteria. Evolutionary pressures differ from habitat to habitat, which possibly results in different functional and structural requirements and consequently a considerable variation in rate of change in different phylogenetic lineages of a protein. As the phylogenetic distances between taxa is at least an order of magnitude greater than between eukaryotic taxa, it may be argued that the informational content is greatly diminished in the comparatively short cytochrome c sequences. Two different scenarios can be envisaged:

-Firstly, assuming a conservation of physicochemical properties and tight structural/functional constraints, the closing in on the limit of change prescribed by structure/function would result in an increasing incidence of convergent mutational events. Ultimately an equilibrium of dissimilarity would be reached. Given enough time, proteins from a range of species would become equidistant from one another (Meyer *et al.*,1986). In other words, they have lost their phylogenetic attributes and the sequence itself may be considered in part "phenetic". Meyer *et al.* argue

that this is the case for cytochromes c2 in prokaryotes and furthermore, that this might also be the case for rRNA sequences, making an all inclusive phylogenetic trees for prokaryotes based on these sequences doubtful.

-Secondly, the limit of change in a particular protein structure may be so relaxed, that the proteins might diverge to such an extent that no degree of sequence similarity can be found. Divergence to such degree would be enhanced by any change in functional role of the proteins. Digital analysis of such sequences would be impossible and the only indication of a common origin of two such sequences would be their three dimensional structures.

In the light of the criteria for an ideal macromolecule as a phylogentic chronometer, recounted in section 1.1.2, cytochrome c is in many aspects not suitable for covering relationships in the procaryote world.

The distribution of cytochromes c is sporadic. It does not occur in many bacterial groups and even in the same genus the distribution of structurally homologous proteins can be erratic (Ambler, 1985). However, the existence of the corresponding genes (Daldal *et al.*, 1986) cannot be ruled out. It could be silent for some reason or another, or expressed in such minute amounts that its product is virtually undetectable. Also, the question of paralogy as opposed to orthology arises. Compared to mitochondria, which has only two types of cytochromes c, the small soluble cytochrome c and the larger membrane bound c, there is a far greater variety of c-type cytochromes in prokaryotes. *Pseudomonas aeuriginosa*, for example, has at least five different cytochromes (Ambler, 1982). Bacterial cytochromes c differ in many physicochemical aspects such as redox potentials, isoelectrical points and details of their spectra, and sequencing has revealed that there are at least four unrelated c-type cytochrome classes. Of the three major structural classes of cytochromes c, characterised from bacteria, Class I comprises by far the largest number of sequences today. Cytochromes c in this class are homologous to the mitochondrial cytochrome c, they are however, subdivided into structurally different subclasses and the lineage is not obvious. Furthermore, the function and structure of homologous cytochrome c, varies to a greater or lesser degree between species. This casts doubt on the assumption that the mutational rate is the same for the cytochromes c as they could be, under different selective pressures or have different size and composition of covarions and therefore different limits of change. It is also apparent that the phylogenetic distances are in an order of

magnitude greater between taxonomic units of bacteria than between eukaryotic taxonomic units. Therefore the mutational rate is, perhaps not, sufficiently slow for cytochromes c to cover the deepest branches of the phylogenetic tree, even if orthologous cytochromes are identified in the different lines of descent. Finally, an important factor in choosing cytochromes c for phylogenetic studies is the relative ease by which they can be obtained, but this has not always been the case. Elaborate methods may have to be used and the quantity of a particular cytochrome c in certain bacteria is often so slight that a sufficient amount is often difficult or even impossible to obtain. However, given that the amount of protein is sufficient, processing the informational content is not a major problem. The use of automated sequenators allows for sequencing picograms of protein and with ever increasing rapidity. Partial sequencing of such proteins can also allow for the possibility of making oligonucleotide probes to make feasible the isolation of the gene itself. On the other hand, the use of heterologous probes cannot be relied upon for bacterial cytochromes, even between species of the same genus, as sequence difference is so great.

Cytochromes c have mainly been isolated from purple bacteria. Trees based on the different paralogous cytochromes in this group exhibit some incongruence to each other as well to dendrograms based on 16S rRNA analyses. This is mainly manifested in the length of branches within taxonomic groups, which can be explained either by parallelism or mutation saturation in the cytochrome lineages, but the more severe example of radically different branching order in the cytochrome c′ and the 16S rRNA tree for *Rhodocyclus tenue* and *Chromatium vinosum* would be best explained by lateral transfer of cytochrome c' (Ambler, 1985; Moore and Pettigrew, 1990).

The 16S rRNA tree and the cytochrome c2 figure tree have the same general topology when insertions and deletions are discounted, but differ more considerably when they are taken into account (Ambler, 1985). Pettigrew and Moore (1987) propose, that where the trees differ, the 16S rRNA tree is the more reliable. Firstly, cytochromes may have undergone functional adaptations and periods of rapid change while 16S rRNA have remained functionally conserved and, secondly because the cytochrome tree shows the effect of mutational saturation at remote divergence, which the 16S rRNA tree does not.

Despite limitations, comparisons of cytochrome c sequences (and of protein sequences in general) are highly informative. Cytochromes c are

found in representatives of many phylogenetic groups, which may be considered important from an evolutionary point of view. They are found in the purple bacteria, Gram positive bacteria, and in *Thermus thermophilus*, of an early branching phylum in the eubacterial tree (the phylum of Thermus/Deinococci, Woese, 1987). Spectral evidence for their presence is also found in archaebacteria in the genus of *Halobacterium.* (Lanyi, 1968). Analyses of cytochrome c sequences may therefore provide a necessary counterpoint to phylogenetic analysis based on 16S rRNA sequences, highlighting specific problems, and clarifying relationships difficult to account for by other means. Cytochrome sequences, as well as those of other electron transport proteins, are also of an obvious importance for relating the evolutionary history of the metabolic machinery to phylogeny based on 16S rRNA analysis.

From a biochemical point of view, the many and varied cytochrome sequences already obtained, pose the challenge of elucidating the evolutionary history of the protein and its many adaptations. The problems, as recounted above, may at present seem insurmountable but it may prove a feasible if not a formidable task with greater knowledge of the evolutionary process in conjunction with secondary and tertiary structure analyses. However, additional sequence information is needed and from a greater range of phylogenetic groups than is presently available. If lateral gene transfer has not played too great a part in the distribution of c-type cytochromes the phylogenetic framework provided by 16S rRNA analyses will, in this context, be invaluable.

1.1.5. Bacterial Phylogeny and Ribosomal RNA

If analysis of cytochrome c sequences can be said to have laid the foundation for phylogenetic studies by establishing in biology the concept of molecular sequence clock, the analysis of rRNA studies has taken the studies a step further, apparently extending the tree to cover the whole prokaryotic world.

The reasons for the success of ribosmal RNA analyses for phylogenetic explorations are the following: 16S rRNA constitute a significant component of cellular mass and are generally easily recovered from all types of organisms. They are ancient and universally distributed. They are extremely conservative in overall structure and the conservation extends to the nucleotide level. It can also be argued that the concept of molecular

clock may be applied to 16S rRNA to a greater degree than to metabolic gene and protein sequences, for the following reasons.

1) 16S rRNA structures as functional entities may be evolutionary inflexible because of strict base bonding requirements and relative rigidity of secondary structures as compared with proteins. They may therefore be inherently more conservative in primary structure.

2) Ribosomes are fundamental for expression of all genomic information. Therefore a functional consistency is required.

3) Changes in environment (with the possible exception of temperature) are not likely to have direct effect on the protein translating apparatus.

Woese (1987) points out that although the greater part of the molecule is highly invariable, short regions vary to a greater extent between species and taxa. This can then be explained by different functional and structural constraints on different segments of the molecule. Consequently, 16S rRNA will allow for measurements of both large and short evolutionary distances.

There are some features which can distort the phylogenetic reconstruction from 16S rRNA sequence data. Structural differences have been observed among bacteria, possibly reflecting different functional demands. Also the G+C content of genomic DNA is diversified to a considerable extent in bacteria, ranging from 25% -75%, suggesting that some selection pressure has operated to alter the G-C content of the genome (Hori and Osawa, 1986). This pressure could have affected the rate of nucleotide substitution. Finally, although 16S rRNA has not yet been shown to be subject to lateral transfer, the possibility cannot be excluded. An interesting possibility might be interspecies recombination conferring antibiotic resistance to the species involved. In many instances antibiotic resistance/sensitivity is specified by the 16S rRNA (Moazed and Noller, 1987)and has been shown to be group specific. Selection for antibiotical resistance has clearly been a strong selective factor for lateral transfer.

1.1.5.1. Phylogenetic reconstruction based on 16S rRNA analyses

The early oligonucleotide catalogue analysis of 16S rRNA analysis used T1 oligonucleotide catalogues of 16S rRNA which consisted of fragments from 1 to 20 bases. Catalogues were then compared, using the SB similarity coefficient. SAB = NAB/2(NA+NB), where NAB is the number of common nucleotides of fragments, five nucleotides and bigger, and NA + NB is the total number of nucleotides. The resultant pairwise distances were then, by cluster analysis cast into dendrograms or trees.

Among the successes of the oligonucleotide analysis was a substantial phylogenetic grouping at various higher taxonomic levels and also the determination of branching order of taxa at the lower levels. The most unexpected result was the discovery of the three archaebacterial, eubacterial and eukaryotic urkingdoms (which has since become a matter of some debate (Lake, 1988). Also, the taxonomic distinction, made between nonphotosynthetic and photosynthetic purple bacteria, was shown to be unfounded phylogenetically (Stackebrandt and Woese, 1981; Woese 1985; Stackebrandt, 1985). This has been supported by sequencing of cytochromes c from representative species.

The phylogenetic subgrouping of bacteria, on the basis of oligonucleotide catalogue analysis, can be validated by so called signature analysis. Some positions in the 16S rRNA molecule are highly invariant and are observed to be characteristic (the signature) of the different phylogenetic units. Furthermore, bases in common to all or most phyla in signature positions, define the ancestral signature for each urkingdom and changes from this ancestral signature in a phylogenetic unit within a particular phylum can be used as a measure of mutation rate. Fast mutating (or evolving) groups are expected to have a greater proportion of the ancestral signature positions changed than slower evolving groups. It is therefore possible to distinguish between truly ancient deep branching species and fast evolving deep branching species. However, the evolutionary branching order is best accomplished by finding the root by including outgroup species to the phylogenetic unit under analysis. This is only possible by analysis of complete sequences when covering the greater phylogenetic distances. Still signature analysis revealed that bacteria separated into more or less naturally defined phyla which was not apparent from the SAB values of oligonucleotide catalogue analyses (Woese, 1987).

Despite considerable successes it was clear that oligonucleotide catalogue analysis was inadequate in many respects, especially in attempts to elucidate the branching order in taxa including rapidly evolving lines of descent, as well as resolving the branching order of distant phylogenetic groups (Woese, 1987).

Now, the oligonucleotide catalogue approach has largely been superseded by complete 16S rRNA sequencing, which makes the full information content of the molecule available for phylogenetic analysis. Hence, the resolution has been increased considerably and a deduction of evolutionary branching order at the greatest phylogenetic distances has been attempted. The most comprehensive and detailed analysis has been made by Woese (1987) with an augmented distance matrix method. It has confirmed the apparent grouping of bacteria into the phyla demarcated by signature analysis and furthermore revealed that the point of divergence of these phyla is at, approximately, the same point in the tree. It would be tempting to relate this explosive radiation of eubacteria to some major change in the abiotic environment such as marking the beginning of the oxygen atmosphere or the advent of photosynthesis, but so far no comprehensive hypothesis embracing this has been presented.

As has been recounted above, the feasibility of an accurate placement of the internodes in the region of this major eubacterial divergence has recently been criticised by Sneath (1989) on the basis of sample error effect (see section 1.1.2.1.1). Thus, even if the overall topology has been defined, the finer details may still have to be worked out.

The branching order of the most distantly related bacteria that define the three urkingdoms has been criticised by Lake (1988). Lake maintains that the distance matrix method is very sensitive to systematic errors due to unequal rates and this has resulted in the wrong tree to be chosen. He proposes an alternative tree, which differs in fundamental aspects. The tree is deduced by the method of evolutionary parsimony, which he claims is more robust in dealing with unequal rates in 16S rRNA lineages. It takes into account that transitions are more common than transversions and reduces the weight of the former to minimise the effect of homoplastic mutations. By this method, *Halobacteria* and *Methanobacteria* cluster with the eubacteria and eukaryotes with the remaining archaebacteria. This has been supported phylogenetic analysis of DNA polymerases by Sidow and Wilson (1990). They have developed the method of evolutionary parsimony further, to take into account the effect of the characteristic

overall base composition of the species under study. This method is claimed to be even more robust than the original evolutionary parsimony method.

1.1.5.2 16S rRNA phylogeny versus evolutionary history of metabolic pathways

One method of unravelling the evolutionary history of metabolic pathways is by inference from the 16rRNA phylogenetic tree, on the basis of distribution of phenotypic characteristics. The assumption is made that, if a complicated trait is found in widely divergent species it must have been present in the common ancestor. In evolutionary studies of bacteria, the evolution of the metabolic machinery must be of central interest, as it relates to the larger history of the world itself, how the living organism has been shaped by abiotic forces and vice versa. Subsequently the question arises, to what extent does the 16S rRNA tree describe the evolutionary history of functionally different biochemical pathways or even the bacterium as a whole.

In principle, the 16S rRNA tree is only the evolutionary history of the protein synthesising apparatus. For example, phenotypic traits encountered in extant bacteria can have an independent origin and be more or less unrelated on the molecular level. Also, different routes can possibly be mapped out for different traits because of the occurrence of lateral transfer. At this stage, neither the fraction of functions subject to lateral interspecies gene transfer is known, nor the frequency of such events relative to geological time scale (Krawiec and Riley, 1990).

While it is clear that the 16S rRNA analyses and subsequent phenotypic mapping onto to the resultant tree is important for the reconstruction of the history of the metabolic machinary, especially when other information is not available, it is at least in principle limited. Sequencing of genes and proteins central to metabolic pathways, in conjunction with structural and functional analysis, is of an obvious importance, both from a biochemical and organismal point of view. It enables the tracing of the origins and the evolutionary routes of the various specialised metabolic pathways as it is the only method that can distinguish between true and superficial resemblance. Sequencing of genes and proteins may provide evidence for, or against, lateral transfer on the basis of comparison between different trees and it will supply answers to questions, as to what constitutes a complex electron transport pathway. Are the various metabolic phenotypes, due to simple independent modifications to a basic electron transport apparatus, shared by most eubacteria; or do they

represent specialised units of many integrated components, unlikely to have evolved but once? 16S rRNA phylogenies can be considered to provide hypotheses to be tested and which will either be strengthened or rejected on the basis of independent evidence.

1.1.6. The Evolution of Metabolic Pathways

The biochemical evolution of the metabolic machinery in eubacteria is generally considered to have followed the sequence:

Anaerobic fermentation ------> Photosynthesis ------> Respiration

ATP, NADH and ferredoxins are considered to be among the earliest constituents of energy metabolism stemming from an ancestral anaerobic and fermentative existence. During evolution from an anaerobic to an aerobic mode of living, other constituents of the metabolic machinery were sequentially introduced, the most pivotal being the various types of electron carriers of increasing redox potentials. The occurrence and evolution of these components is thought to have been directed by the steadily increasing oxidation state of the environment and the availability of nutrients and minerals (Jones, 1985). According to this scenario, cytochromes b are considered to be pre-photosynthetic, but cytochromes c are thought to have arisen during the evolution of photosynthesis (Jones, 1985). They had high redox potentials and increased the efficiency of this particular pathway in an increasingly oxidised world. Consequently they played an important part when photosynthetic machinery was adapted to respiration. That change was steered by further oxidation of the environment and the accumulation of minerals suitable as terminal oxidants, the ultimate one being oxygen (Dickerson, 1980; Jones, 1985).

The order of events presented above, is supported by physicochemical analyses of the various components in different metabolic pathways, and is backed, by sequence comparisons of ferredoxins from anaerobic fermenting and photosynthetic bacteria, and cytochromes c from various Gram negative photosynthetic and aerobic bacteria (Dayhoff *et al.*, 1978). According to this analysis, the ancestral fermenting bacteria are represented in extant bacteria by the fermenting clostridia. This has been contested by Woese and Stackebrandt (1985) on the basis of 16S rRNA phylogenetic analysis of these and related species. Photosynthesis is found

in most eubacterial phyla, including the phylum of Gram positive bacteria and it is therefore postulated that photosynthesis is ancestral to these lineages and even that it is the oldest eubacterial metabolic trait. The phylogenetically deep group of fermenting anaerobic bacteria in the phylum of Gram positive bacteria could then be explained by an early selective loss of the major part of the photosynthetic apparatus.

16S rRNA analysis also suggests that aerobic respiration has evolved independently in the different phyla and also independently within a phylum. This is based on the observation that strictly anaerobic species are found to be members of deep phylogenetic clusters whereas related aerobic species belong to relatively shallow clusters.

In figure 1.1 is presented a schematic figure by Moore and Pettigrew (1987) based on present theories about the evolutionary course of the energy metabolism. It is related to extant eubacteria, in the phylogenetic order derived from 16S rRNA analyses, and the distribution of cytochromes c. It is proposed that the origin is the appearance of photosynthesis and that the ancestral phototrophs used H_2S as a electron donor. These diverged to give rise to purple bacteria and green bacteria. *Rhodospirillaceae* evolved the ability to use organic molecules as electron donors and the cyanobacteria developed oxygenic photosynthesis using H_2O as electron donor. It is also proposed that aerobiosis arose independently in the different eubacterial lineages. The α-subdivision of aerobes are supposed to have diverged from *Rhodospirillaceae* and the β-lineage from an early purple bacterial ancestor, The γ-group aerobes are proposed to stem from the *Chromatium* lineage.

In *Rhodospirillacea*, it has indeed been shown that the organisms use the same electron transport chain for aerobic respiration and anaerobic photosynthesis (Scherer, 1990). It can therefore be hypothesised that the evolution of respiration is due to different and independent modifications of the photosynthetic pathway in the various phyla, and different electron transport proteins are to be found in the independent adaptations. This, however, remains to be conclusively proven and the latest sequence information may give evidence to the contrary (see discussion, chapter 6).

Woese (1987) minimises the contribution from lateral gene transfer in the distribution of metabolic traits, and points at the congruence of the cytochrome c2 tree and 16S rRNA tree for the alpha subgroup of purple bacteria. An alternative scenario, favoured by some authors, is that lateral gene transfer has played a major role at certain stages in the evolution of

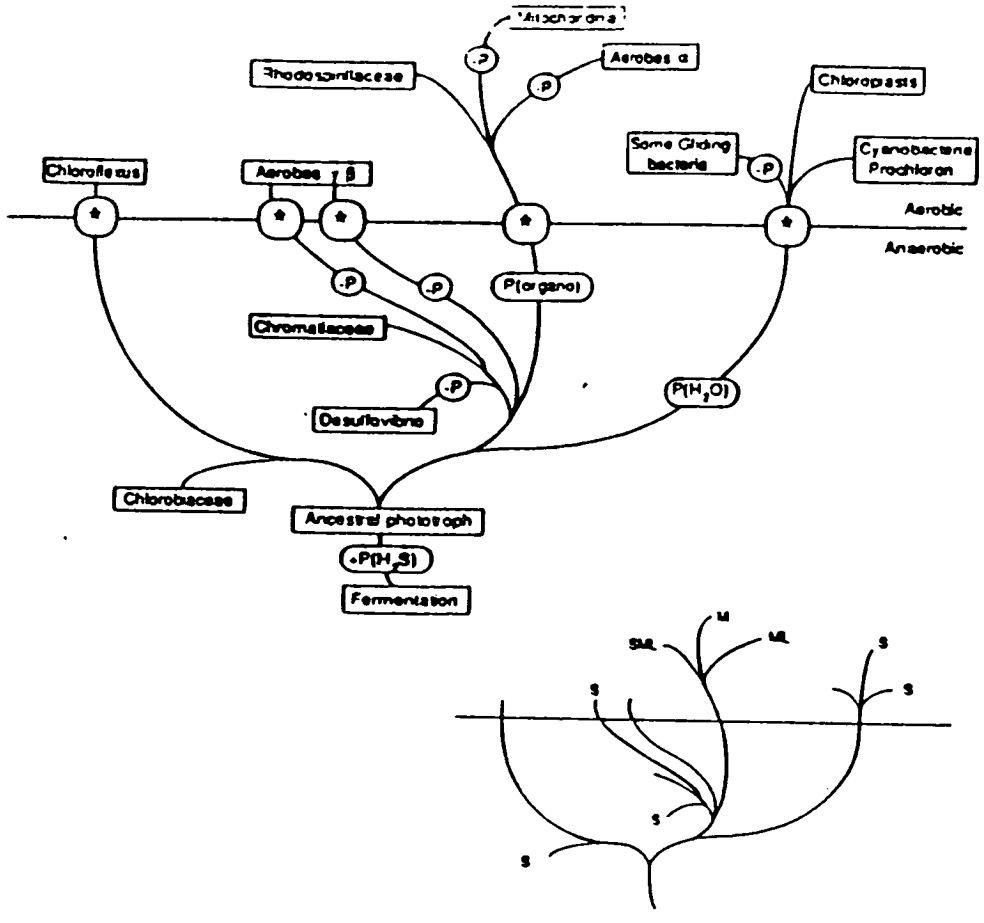


Figure 1.1. The evolution of energy metabolism.
This figure is a schematic representation of the evolution of energy metabolism and how it relates to extant bacteria by Moore and Pettigrew (1990) . The origin is proposed to be the appearance of photosynthes. It is hypothesised that the ancestral phototroph used H_2S as electron donor and diverged to give purple and green bacteria. The Rhodospirillaceae developed the ability to use organic molecules as electron donors (P(organo)) and the cyanobacteria evolved oxygenic photosynthesis, using water as electron donor (P(H2O)). The increasing concentration of oxygen led to the multiple appearance of aerobic respiration (*) and the selective loss of photosynthesis in some lineages (-P). The branching pattern follows the pattern of 16sRNA sequence analysis. This places Chlorobiaceae as an early divergence and groups *Desulfovibrio* with *Chromatium*. The α-group of aerobes (Woese et al., 1985) derive from the Rhodospirillaceae, the β-lineage from an early purple bacterial ancestor, and the γ-group from the *Chromatium* lineage. The occurrence of small (S), medium (M) and large (L) (Dickerson, 1980, table 6.3) cytochromes c in the evolution pattern is shown in the lower diagram.

metabolic systems in various groups. Ambler (1991) suggests that this is the most plausible explanation for the sporadic occurrence of the same complex metabolic functions such as denitrification, methylotrophy, nitrogen fixation besides aerobiosis and photosynthesis, in most of the major eubacterial phyla. If these functions have been transmitted vertically, the primitive organism had the capability of carrying out all these functions.

Possible indicators of lateral transfer of cytochromes c may be the similarity of c'-type cytochromes from *Rhodocyclus tenue* and *Chromatium vinosum* as previously mentioned, also part of genes may have been transferred as appears to be the case between the cytochromes c2 from *Paracoccus denitrificans* and *Rhodopseudomonas capsulata,* the N-terminal one third has 89% identity, whereas for the C-terminal two-thirds, there is only 36% identity (Ambler, 1985). Furthermore, the correlation of "*Pseudomonas aeruginosa* cytochrome c551" with the sporadic ability to denitrify, may suggest lateral transfer.

Electron transport proteins at the low-free-energy end of the electron transport chain may be expected to be peripheral to the main body of the electron transport chain, and able to confer a clear selective advantage to anaerobic photosynthetic species with the same core of electron transport proteins. Occasionally integrated metabolic systems may also have been laterally transferred, as seems to be the case for nitrogen fixation genes, leading to a complete reconstruction of more primitive pathways.

Meyer (1991) has proposed that lateral gene transfer has played a part in the evolution and distribution of the photosynthetic machinery. This is based on the structural and biophysical properties of the reaction centres and similarities between cytochromes c.

It can be concluded that while there can be a consensus about the biochemical order in which the components of the metabolic pathway have evolved, there is a considerable difference of opinion in how it relates to the overall evolution of bacteria or the bacterium as an entity. Both of the alternative theories depend largely on phenotypic analyses. The first one supports a cointegrative vertical evolution of biochemical traits with lateral transfer playing a minor role, and makes the perhaps unwarranted assumption that the evolutionary history of biochemical pathways can be inferred by the mapping of complex phenotypes onto the 16S rRNA tree when relevant genotypic data is missing. The questions that need to be answered are: Are these traits indeed complex and monophyletic and, also, has the correct tree been chosen? Even if the

methods currently most commonly used (distance matrix and parsimony) of phylogenetic reconstruction are mostly right, they can be wrong in places, and one will not know where. (Sidow and Wilson, 1990). The other hypothesis proposes a more disruptive network like evolution, which would be inherently more difficult to decipher. This remains to be proven, and unexpected similarities can have different explanations such as convergence parallel evolution, mutation saturation due to functional and structural restraints and recruitment of the same gene/protein to the same function in different lines of descent.

1.2. PHYLOGENY OF GRAM POSITIVE BACTERIA

The basis of the original definition of this group was the positive Gram stain response, which reflects a very distinct cell wall structure. The Gram positive bacteria were also found to lack an outer membrane and therefore they do not have a periplasmic space. Many species have the ability to form heat-resistant spores. The fact that these phenotypic traits were valid phylogenetic markers was later born out by 16S rRNA oligo-catalogue analyses. The ability to form spore has not been found outside the phylum and only one other phyla the *Deinococci* gives the Gram positive staining response. The Gram positive bacteria form a tight phylogenetic cluster, which includes, unexpectedly, also some Gram negative staining species, of which some are able to form spores (Woese, 1987).

Apart from a few complete 16S rRNA sequences, the classification of bacteria in the phylum of Gram positive bacteria, is mainly based on oligonucleotide catalogue reported in Stackebrandt and Woese (1985) and the revisions made on the basis of complete 16S rRNA sequence analysis reported in Woese (1987). Although incomplete, these analyses are the most exhaustive to date and provide the necessary phylogenetic framework for any attempts to trace the twists and turns in the evolutionary path of the phylum.

On the basis 16S rRNA analyses, the Gram positive phylum is subdivided into four major groups. The two major and best characterised groups are the clostridial and actinomycetal subdivisions, readily distinguished on the basis of DNA composition. The former has a low G+C percentage, (<50%), and the latter high G+C DNA (>55%). The sequence signature of the

clostridial subdivision conforms more to the ancestral eubacterial signature and is therefore considered to be the more ancient group.

The clostridial subdivision is mainly anaerobic, spore-forming and rodshaped. However, one or more of these traits have been lost in several lines of descent. It consists of at least five major sublines, including the metabolically diverse *Bacillus* subline, which apart from the mostly aerobic genera of *Bacillus* and *Staphylococcus*, contains strictly anaerobic species such as *Clostridium inoccuoum*, and microaerophilic genera such as *Lactobacillus*, *Streptococcus* and *Mycoplasma* (Woese 1985, 1987). According to oligonucleotide catalogue analysis, this subline is relatively shallow. The remaining four sublines are strictly anaerobic and most of them contain clostridia.

The actinomycetal subdivision is relatively shallow and the subgroups are mostly aerobic except for the two deepest branching groups the *Bifidobacteria* and the *Propionibacteria*.

The third subdivision consists so far of only one species, the strictly anaerobic photosynthetic *Heliobacterium chlorum*, which is Gram negative staining. However biochemical analysis of the cell wall, indicates Gram positive cell characteristics (Gest, personal communication). The fourth subdivision is represented by the Gram negative staining species of *Selenomonas*, *Megaspaera* and *Sporomusa* (Woese, 1987).

The definition of the last two groups seems to be largely based on their distinct phenotypic character, as their 16S rRNA sequences are definitely of the clostridial type (Stackebrandt, 1985; Woese, 1985; Woese 1987).

1.3. BIOCHEMISTRY OF CYTOCHROMES C

Cytochromes c are defined as cytochromes with covalently bound haem groups. In cytochromes c characterised so far, side-chains of the haem are connected with the protein by two thioesterlinks to two cysteine residues, separated by two (or in rare cases four) amino acids. Exceptions are found sush as in the cytochromes c from the mitochondria of some protozoa, where the haem is linked to the protein chain by only a single thioesterlink (Pettigrew, 1974).

In the early stages of biochemistry, it was recognised that proteins were species specific. Later it was shown that this specificity is quantitatively associated with a particular molecular structure. As it was realised that the specificity of proteins is ultimately determined by their

amino acid sequences it became evident that the determining of primary structure would be of fundamental importance for biochemical studies. On the basis of sequence comparisons, Ambler (1982) has defined four unrelated classes of cytochromes c: Class I, Class II, Class III and Class IV. Of these classes I, II and III are the best characterised. Their existence is further supported by X-ray studies that show a characteristic fold for the proteins in each class.

Class I comprises the most sequences today and, on the basis of primary structural features, it has been further divided into five subclasses. Three dimensional structures of cytochromes of class I, derived from X-ray crystallographical studies have revealed that they are small compact proteins. The protein chain is wrapped tightly around the haem, leaving little room for alpha- and beta-configurations prominent in other proteins. The haem is buried in such a way that two polar side chains are under the surface of the protein. This is not the most favourable position from an energy standpoint of view and therefore the side-chains are stabilised in this position by hydrogen bonding. Consequently, only one edge of the haem is exposed, which seems to be a prerequisite for a proper flow of electrons in and out (Dickerson, 1980; Salemme, 1977). Differences in redox potentials between the various c-type cytochromes are probably mediated by the degree of shielding of the haem, exposure of the propionates and subtle changes in the region surrounding the haem (Pettigrew and Moore, 1987; Moore and Pettigrew, 1990).

On the basis of sequence comparisons, regions can be defined according to the degree of variability in the amino acid sequence. Such comparisons of mitochondrial cytochromes c have for example shown that the region around the haem crevice is very well conserved. Other amino acids in specific positions are shown to be completely invariant, indicating an important structural role and some regions or positions, have a specific types of amino acids, according to where they are in the folded protein. Hydrophilic amino acids are on the outside and hydrophobic ones in the interior. A certain amount of change is allowed in these regions as long as amino acid of the similar chemical character is substituted for the ancestral one (Dickerson, 1972; Salemme, 1977).

In extant organisms, cytochromes c of Class I are thought to play many different roles. Their common feature is their small size and a high redox potential and they are thought to have a similar role as electron carriers near the low-free-energy end of photosynthesis and respiratory electron-transport chains. Most of the ones isolated until now, are water soluble and

are thought to shuttle along the outer membrane between a preceding reductase and various electron acceptors. Meyer and Kamen(1982) have proposed at least three roles for cytochrome c2:

(1) as an immediate electron donor to reaction centre chlorophyll in those species which do not have membrane -bound cytochrome c556 c552;
(2) as an electron donor to a bacterial cytochrome oxidase and ultimately oxygen in the facultative aerobes, and
(3) as an electron donor to a cytochrome cd1 type of nitrite reductase.

This could account for outer regional differences between the various cytochromes in type and distribution of charged amino acids (Dickerson, 1971), the more variant regions being at sites in the cytochrome contacting the different types of oxidases or photoreaction centres. Kamen *et al.* (1977) have pointed out that bacterial cytochromes c generally exhibit good reactivity with mitochondrial reductase, but poor reactivity with mitochondrial oxidase. This was interpreted such that the reductase was a common evolutionary heritage, whereas the oxidases are more diverse. This appears to support the hypothesis of the independent evolution of aerobiosis in different lines of descent in the all inclusive phylogenetic tree.

1.4. PROJECT DESCRIPTION

1.4.1. Aims

This thesis presents work, which was undertaken to obtain sequence information on cytochromes c from *Bacillus* species, the examination of the informational content of class I cytochromes c and their significance for the elucidation of the evolutionary history of eubacterial energy metabolism, with a special emphasis on Gram positive bacteria.

1) Cytochrome classification or taxonomy.
Any sequence will only be interpreted by its context. The bases of cytochrome c classification will be examined and attempts will be made to evaluate the affinities of currently unclassified cytochromes to those more thoroughly studied. Attempts will be made to expand the current cytochrome c classification scheme of Ambler (1977) to incorporate

cytochrome c sequences from *Bacillus* obtained in this work or from other sources. It is expected that some information about the nature and diversity of cytochromes in the genus of *Bacillus* will be gained. In the light of the phylogenetic distance between the Gram negative bacteria and Gram positive bacteria it will be interesting from an evolutionary vantage point to see if cytochromes c from *Bacillus* fall into any of the already defined groups of cytochromes from Gram negative bacteria or if they make their own.

2) Investigation of the relationship of the electron transport chains of Gram positive bacteria to electron transport chains of bacteria in other phyla.

The anaerobic substructure of the phylum of Gram positive bacteria has been noted by various authors such as Dayhoff *et al.*,(1978) and Woese (1987). The latter has proposed that aerobiosis has evolved independently in the phylum of Gram positive bacteria. Anlyses of electron transport protein sequences from aerobic species belonging to the genus *Bacillus* may shed light on the origin of aerobiosis in the phylum.

3) The roles and cellular location of cytochromes c in the genus *Bacillus*.

While extensive physiological studies will not be undertaken, the structural and biochemical properties of cytochromes c may indicate their role in the metabolic spectrum of the species and adaptations to different habitats. Cytochromes may be induced in the presence of different metabolites. electron donors or terminal electron acceptors, such as oxygen or nitrate and conversely others may be repressed. Furthermore it has been observed that certain metabolic processes and particular structural types of cytochromes have coincident distribution.

In Gram negative species, it is now generally accepted that cytochromes c are located outside the cytoplasmic membrane in the periplasmic space. Many of these cytochromes are found to be soluble entities. The placement of cytochromes outside the cytoplasmic membrane is considered a crucial feature of electrogenic transfer to a terminal oxidant site at the cytoplasmic side of the membrane (Pettigrew and Moore, 1987). Gram positive bacteria do not have an outer membrane. If the roles and functional requirements of cytochromes in Gram positive bacteria are the same as in Gram negative bacteria, specific adaptations must have evolved to retain the cytochromes at the outer surface of the cytoplasmic

membrane. This may be studied indirectly by spectrophotometric and elecrophoretic analyses, but in conjunction with analyses of gene and complete protein sequences the mechanisms of attachment should be revealed.

The aims of the experimental work were.

1) The isolation and cloning of the gene for the cytochrome c552 from *Bacillus licheniformis* for structural and evolutionary analyses and to obtain sequence information on its mode of attachment to the outer surface of the cytoplasmic membrane.

Partial sequences of the cytochrome c had already been obtained by Woolley (1984) in this laboratory. The sequence information obtained allowed oligonucleotide probes to made from a region of low redundancy, which covered the haem site.

2) The examination of the cytochrome c content in *Bacillus azotoformans* under aerobic, and denitrifying anaerobic conditions, the investigation of the cellular location of cytochromes, and the characterisation and purification of cytochromes c from for sequencing, and subsequent structural and evolutionary analyses.

Simultaneously the protein sequencing of the cytochrome c552 from *B. licheniformis* was continued in a collaborating laboratory by vanBeeumen. During the course of my work he obtained the complete haem domain sequence. The sequence is presented in this thesis and discussed along with the sequence information obtained on the cytochromes c from *Bacillus azotoformans.* A cytochrome c sequence from *Bacillus subtilis* was also reported during the course of the work (Von Wachenfeldt and Hedterstedt, 1990) and Sone and coworkers which have done impressive work on the electron transport chain in a thermophilic strain *Bacillus*PS 3 have reported the gene sequence of a cytochrome c apparently fused to the aa3 cytochrome c oxidase complex (Ishizuka *et al.*, 1990) . They have also isolated and sequenced the gene of a membrane bound cytochrome c and a paper on the results (Fujiwara *et al.*, 1993), to which they have kindly given me access to, has been submitted to Biochimica and Biophysica Acta. These sequences will also be classified

and made use of in the discussion of the evolution of the electron transport chain in the genus *Bacillus* and the energy metabolism of Gram positive bacteria in the context of eubacterial phylogeny.

1.4.2. Bacteria Under Study

Bacillus licheniformis is one of the oldest denitrifying species known and was described as early as 1910 by Beijerinck and Minkman (Pichinoty, 1978). It is Gram positive endosporeforming rod, a facultative anaerobe, capable of fermentation as well as of nitrate and oxygen respiration (Claus and Berkeley, 1986). It is very similar to *Bacillus subtilis* in physiological and morphological characteristics, and many strains were formerly classified as *B. subtilis*. On examination of the phenotypic data on these bacteria presented in Bergey's Manual a distinction can be made between *B. subtilis* and *B. licheniformis*, mainly on the basis of the ability of the former to reduce nitrate to nitrite, and tolerance to potassium (Claus and Berkeley, 1986).

Studies on electron transport chain in *B. licheniformis* focus mainly on its ability to reduce nitrate. *B. licheniformis* has been reported to be capable of denitrifying respiration. Some strains of *Bacillus* were classified under a name, making a distinction of this ability, as *Bacillus denitrificans*. They have now been reclassified under the name of *B. licheniformis* (Pichinoty, 1978).

Bacillus azotoformans is a much "younger" species and consequently much less studied than *Bacillus licheniformis*. Its isolation was described by Pichinoty in 1976. It is a Gram negative rod, motile with peritrichous flagella. It produces oval spores without exosporium in swollen sporangia. However, the walls are thick, it has mesosomes and persistent septa characteristic of true Gram positive bacteria, It lacks fermentative activity and does not degrade polysaccharides.

While both the *Bacillus* species are denitrifiers, *Bacillus azotoformans* is much more vigorous as such than *Bacillus licheniformis*. Pichinoty was unable to detect nitrite reductase in *B. licheniformis*. This is surprising as nitrite reductase is a key enzyme in the denitrifying electron transport pathway. Denaries *et al.*,(1991) have recently isolated the corresponding enzyme from *Bacillus halodenitrificans*. It was observed that the enzyme was very sensitive to oxygen, and had to be purified under anoxic conditions in the presence of argon. If the nitrite reductase is expressed

only in low levels in *B. licheniformis* this may be the explanation for its conspicuous absence..

There are not many taxonomic studies which relate these bacteria to each other. To my knowledge they have only both been included in a study on DNA base composition where they are shown to belong to different taxons (Priest, 1981).

CHAPTER 2
2. MATERIALS AND METHODS

2.1. STRAINS OF BACTERIA, PHAGES AND PLASMIDS

2.1.1. Bacteria Under Study

Bacillus licheniformis 6346: A constitutive penicillinase mutant, derived from NCIB (The National Collection of Marine and Industrial Bacteria) 6346, and designated 6346/c.

Bacillus licheniformis 749*: Strain collection at the ICMB, University of Edinburgh.

Bacillus azotoformans: NCIB 11859. The National Collection of Marine and Industrial Bacteria. Aberdeen, Scotland.

2.1.2. Strains of Bacteria Used for DNA Work

E.coli TG1 Used as a host for M13 phage vectors. From Amersham International plc, Amersham, Buckinghamshire

E.coli NM522 Used as a host for M13 phage vectors. From Professor N.E. Murray, University of Edinburgh.

E.coli NM522* Same as above, but also lambda resistant.

E. coli NM514 Used as a host for λ1149 recombinant library phages. The insertion of a donor DNA into the CI gene of the vector is recognised by a change from a turbid to a clear plaque morphology. From Professor N.E. Murray, University of Edinburgh.

2.1.3. Phages

λ1149:	Immunity insertion vector with a single site for *Hind* III and *Eco*RI. Used to make a library from *B. licheniformis* genomic DNA. From Professor N.E. Murray, University of Edinburgh.
λIC I857	Used as a DNA size marker. From Professor N.E. Murray, University of Edinburgh.
M13mp18:	A blue/white response sequencing vector from Amersham Int., plc., Amersham, Buckinghamshire.
M13mp19:	A blue/white response sequencing vector, for the reverse direction, relative to M13mp18, from Amersham Int., plc., Amersham, Buckinghamshire
Mp L 30	An M13 derived blue/white response sequencing vector, which has a polylinker with a number of restriction sites, including a single *Cla*I site. From Dr. D. Leach, University of Edinburgh.

The M13 derived vectors carry a segment of DNA derived from the *lac* operon in *E.coli*. It codes for the amino terminal fragment of b-galactosidase which can be induced with isopropyl–β -D-galactoside (IPTG). The fragment is capable of intra allelic complementation with a defective form of β-galactosidase encoded by the host. If bacteria, containing these plasmids, are exposed to IPTG, they synthesise both fragments of the enzyme and form blue colonies on a solid medium, containing the chromogenic substrate 5-bromo-4-chloro-3-indolyl-β-galactoside (X-gal). Insertion of foreign DNA into the polyclonal site of the plasmid inactivates the amino-terminal fragment of β-galactosidase and abolishes the complementation. Bacteria, carrying recombinant plasmids, give rise to white colonies, therefore the designation blue/white response.

2.2. MEDIA, BUFFERS, ENZYMES, CHEMICALS AND OTHER MATERIALS

2.2.1. Solid Media

Luria-plates

Luria agar plates (L-plates) were used for the growth of bacteria and contained (w/v): agar (1.7%), Difco bacto tryptone (1.0%), Difco yeast extract,(0.5%); NaCl (1.0%); adjusted to pH 7.2 with NaOH. Antibiotics were added (100µg/ml) when the sterilised agar had cooled to approximately 50°C.

BBL-top agar

BBL-top agar contained (w/v): Baltimore Biological Laboratories (BBL) trypticase (1.0%), NaCl (0.5%), Difco agar (0.65%). BBL-top agar was used to visualise plaques, obtained following the transformation/transfection of *E. coli* with phage DNA, and for the isolation of lambda DNA by the plate lysate method.

2.2.2. Liquid Media

Luria broth(L-broth) was used for the routine growth of *B. licheniformis, B. azotoformans* and *E. coli.* It contained (w/v): Difco bacto tryptone (1.0%); Difco yeast extract, (0.5%); NaCl (1.0%); adjusted to pH 7.2 with NaOH. Large scale cultures of *Bacillus* were also grown on 1% yeast extract (Difco), 0.5% NaCl. The medium was supplemented with 0.5% $NaNO_3$ or 0.5% KNO_3 when *B. licheniformis* and *B. azotoformans* were grown under anaerobic conditions.

2.2.3. Buffers and Solutions

These are specified in the appropriate method sections.

2.2.4. Enzymes

DNAase from bovine pancreas, lysozyme, T4-ligase, Klenow polymerase, T4-polynucleotide kinase. Restriction enzymes: *Hind* III, *Eco* RI, *Aat* II, +

Pst I, *Sal* I, *Pvu* I, *Xho* I, *Bam* HI, *Bgl* II, *Xba* I and *Cla* I.alkaline phosphatase from calf intestine, bovine trypsin, V8-protease, thermolysin, chymotrypsin, carboxypeptidase A, protease K, pronase, subtilisin and pepsin. These enzymes were bought from the following companies or made in the laboratory:'

Amersham Int., plc.,Amersham, Buckinghamshire
Pharmacia Ltd., Central Milton Keynes
Boehringer Corporation (London) Ltd, Lewes, Sussex
New England Biolabs, CP Laboratories, Bishop´s ,Stortford.

2.2.5. Chemicals

Chemicals used in this work were purchased from:

BDH Ltd., Pool, Dorset
Sigma Chemical Company, Pool, Dorset
Boehringer Corporation (London) Ltd, Lewes, Sussex.

2.2.6. Biological kits and other materials.

Nitrocellulose filters for Southern blotting were bought from Schleicher & Schull, GmbH Dassel, Federal Republic of. Germany.
H-bond nylon filters for Southern blotting were bought from Amersham Int., plc., Amersham, Buckinghamshire.
DEAE nitrocellulose filters for size selective DNA isolation were purchased from Schleicher & Schull, GmbH, Dassel, Federal Republic of Germany.
Geneclean kit: Used for DNA extraction from agarose gels. Purchased from Bio 101 Inc., La Jolla, USA.
One PhorAll, an all purpose buffer for restrictions, ligations, and polynucleotide kinase reactions, was purchased from Pharmacia Ltd., Central, Milton Keynes
λ-in vitro packaging kit was bought from Amersham Int., plc., Amersham, Buckinghamshire.
Radioactive nucleotides were bought from Amersham Int., plc., Amersham, Buckinghamshire.

Oligonucleotides, used as gene probes or for priming in DNA sequencing experiments were purchased from Oswell DNA Service, at the University of Edinburgh.
DNA ladder markers from Bethesda Research Laboratories, Paisley, Scotland.

2.3. PROTEIN TECHNIQUES

2.3.1. SDS Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS polyacrylamide electrophoresis gels were run to monitor purity, diversity, and sizes of cytochromes c, isolated from *B. azotoformans.* Samples were boiled for 2 min with sample loading buffer, before loading into wells for vertical electrophoresis, in 1-1.5mm x 14cmx20 cm gels. The gels were run at constant approximately 4-5 V/cm, (80-100 V) overnight. The buffer system of Laemmli (1970) was used, according to the recipes below.

Acrylamide stock solution: 30g acrylamide; 0.8g N,N'-methylene-bis acrylamide; distilled H_2O to 100 ml. Degassed and filtered through Whatman No.1 paper, before storage, in a dark bottle, at 4°C.
Resolving gel buffer: 3.0 M Tris/HCl, pH 8.8
Stacking gel buffer: 0.5 M Tris/HCl, pH 6.8.
Stacking gel: 1.25 ml acrylamide stock; 2.5 ml stacking gel buffer; 0.1ml. 10% (w:v) SDS; 0.5 ml 1.5% ammonium persulfate (AMPS); 5.65 ml distilled H_2O; 15 µl TEMED (N,N,N',N'-Tetramethylenediamine).
10X Reservoir buffer: 0.25 mM Tris; 1.92 M glycine; 1%(w:v) SDS.
Sample loading buffer: 62.5 mM Tris/HCl, pH 6.8; 2%(w:v) SDS; 20%(w:v) sucrose; 0.1% bromophenol blue.

Composition of Resolving Gel Mixtures for Polyacrylamide Gel Electrophoresis:

Components	20%	15%
Acrylamide stock (ml)	20.00	15.00
Resolving gel buffer (ml)	3.75	3.75
10% SDS (ml)	0.30	0.30
1.5% AMPS (ml)	1.50	1.50
Distilled H_2O (ml)	4.45	9.45
TEMED (µl)	15	15

2.3.2. Staining of SDS-Polyacrylamide Gels

2.3.2.1 Protein staining

The gels were stained in 0.1% Coomassie Blue (R) in 4:1:5 methanol/Glacial acetic acid/H_2O. The gels were destained in the same solution, without Coomassie blue.

2.3.2.2. Haem staining

The gel was immersed for 30 min, with constant shaking, in 1.25 mM TMBZ (3,3',5,5'-tetramethylbenzidine) in methanol/0.25 M sodium acetate, pH 5, 30:70 by volume. Then, H_2O_2 was added, to give a concentration of 26 mM, and the gel shaken for further 15 min, after which the staining solution was replaced with fresh propanol/0.25 M sodium acetate, pH 5, 30:70 by volume (Goodhew *et al.*, 1986).

~~solubilised~~

2.3.3. Estimation of Protein Concentration

As membrane proteins were the subject of study in the present work which often required detergents for solubilisation, the method for the estimation of protein concentration, was an adaptation of the standard Lowry method, the so called micro-Lowry (Findlay, 1987). To 200 µ l of the protein solution in detergent were added 1 ml of fresh Lowry reagent (1 ml each of 2.0% sodium potassium tartrate and 1% $CuSO_4$ into 100 ml of 2% Na_2CO_3 in 0.1 1N NaOH). The solution was mixed thoroughly and allowed to

stand for 10 min. Then 10% SDS was added to a final concentration of 1% (mixed thoroughly), and 100 µl of 1:2 diluted Folin Ciocalteau reagent (mixed thoroughly). Absorption was read at 660 nm after 30 min and within 2 hours. Values were compared to a standard curve obtained from measurements of serial dilutions of bovine serum albumin solution (BSA).

2.3.4. Spectra

Spectra were recorded on a Perkin-Elmer 320 spectrophotometer and 1 cm pathlength quartz cuvette was used. Spectra were made of purified cytochromes, in the fully oxidised and fully reduced state. They were dissolved in distilled water, or in 50 mM ammonium acetate buffer, pH 7.0. in sufficient concentration to give an absorbency of 0.5-1.0 absorbency units, for the reduced Soret peak. Cytochromes were fully oxidised after the last purification step, and reduced spectra were obtained by adding few crystals of sodium dithionite.

Difference spectra were made of the alpha peak, at various stages, during purification as means of following the changes in concentration. The ratio, reduced alpha peak absorption/absorption at 280nm, was used as a measure of purity.

2.3.5. Measurement of Midpoint Oxidation-Reduction Potentials

Mid-point oxidation-reduction potential of cytochromes c is calculated by a plotting $\log(\text{cytochrome}_{ox}/\text{cytochrome}_{red})$ versus the standard potential, measured at different points during redox titration. $\text{Log}(\text{cytochrome}_{ox}/\text{cytochrome}_{red})$ is determined from the relative heights of the partially and fully reduced alpha peaks. The potential measurements are made directly, using an Ag/AgCl reference electrode, and the standard potential for AG/AgCl 196 mV, is added to the observed potential. The plot produced should be a straight line with a slope of 60 mV.

The redox titrations of cytochromes were performed in 2.8 ml of 20 mM potassium phosphate buffer, pH 7.0, containing 10 µ l each, of 5 mM solutions of the following redox mediators: Phenazine methosulphate and diaminodurol, 30 µl of 5 mM ferric-ammonium sulphate/10 mM EDTA solution, and a sufficient concentration of cytochrome, to give an alpha oxidised/reduced absorbency change of about 0.1-0.2 units. Anaerobic

conditions were maintained in the cell by bubbling argon through the solution for 20 min before titration, and then constantly throughout the experiment. The cytochromes were titrated with sodium dithionite (20mg/ml) and potassium ferricyanide (20mg/ml). After addition of each small portion, the system was allowed to reach equilibrium before scanning, and the absorbance at one of the isobestic points readjusted to a constant value so as to compensate for any minor changes in absorption, caused by the redox mediators (Pettigrew and Moore, 1987).

2.3.6. Dehaemation of Cytochrome C

A salt-free solution of cytochrome c was freeze-dried in a 25-50 ml round bottle flask. Then, 20-30 mg $HgCl_2$, and 3.5 ml of acidic urea (8 M urea, 0.1 N HCl) were added, and the flask swirled until the content had dissolved. The flask was closed and incubated for 16 hours at 37°C. The mixture was then desalted on a Sephadex G25 (fine) column (1.5cm x 20cm). The colourless apoprotein was eluted by gravity and freeze-dried (Ambler, 1963). The haem bound to the column, but could be released from it, if immediately washed with 50% pyridine. The acidic urea solution was made in the following way: 2.4 g of urea was dissolved in a test tube, and 3 ml of 0.1 N HCl added. The tube was whirlmixed until the salt had dissolved. The final volume was about 5 ml.

2.3.7. Amino Acid Analysis

2.3.7.1. Qualitative amino acid analysis

Qualitative amino acid analysis was only performed on peptides. Approximately 10 nmol of peptides were dried, under vacuum, in a 8mm x 35mm glass tube. 100 µl of 6 M HCl was added, the tube sealed, and incubated at 105°C overnight, after which it was opened and the content dried down under vacuum. The sample was redissolved in 10 µ l 0.1 M NH_3 loaded on Whatman No.1 paper, and the amino acids separated by HVPE at pH 2.0. Samples were run at 5 kV (110 V/cm) for 15-30 min.

Methyl green was used as a marker and amino acid standards were also run. After the paper had dried, it was stained with ninhydrin/collidine (section 2.3.9.6). The amino acids present in the peptide, are distinguished by their relative positions and characteristic colours, soon after the

staining. The amino acids are easily recognised, apart from the hydrophobic amino acids which run in almost the same position. Serine also runs into that same position, but can in most cases be distinguished from the others by its characteristic green olive colour. Methionine is only seen as the oxidised product. Tryptophan does not show up readily as it is destroyed in the hydrolysis (Ambler and Wynn, 1973).

2.3.7.2. Quantitative amino acid analysis

Amino acid analysis of peptides and proteins were performed on a Beckman 120C amino acid analyser. It has two columns, one for the detection of basic amino acids and another for neutral and acidic amino acids. Samples, (5nm-10nm for proteins, and 20-30nm for peptides) were dried down under vacuum in 10mm x 125mm Pyrex test tubes, and then dissolved in 0.5 ml of freshly diluted 6 M HCl. A constriction was formed in the tube and the contents frozen. The tubes were then evacuated while the acid thawed. They were sealed and heated at 105° for 24 hours. After being opened, the acid was removed by evaporation, under vacuum, and re-dissolved in 0.35 ml of citrate buffer, pH 2.2, containing known amounts of norleucine and a-amino-g-guanido propionic acid as internal standards, and 150 µl applied to each column. Approximate sample loadings were 5-10nmol protein/column and 10-30nmol peptide/column. Proteins and peptides were oxidised when necessary, to facilitate the quantification of cystine. Proteins and peptides were oxidised, by dissolving them in 100 µ l performic acid (formic acid/30% hydrogen peroxide, 19:1) in screw cap tubes, which were then incubated at 4°C for 2 hours. They were then dried down and repeatedly dissolved in 200 µl and dried down. Sometimes, peptides were oxidised immediately after the HVPE purification step while still in the paper. The dried paper strips containing the peptides,were put in a desicator containing performic acid, which was then partially evacuated.

The detection and quantification of tryptophan was by a separtate experiment, according to the follwing procedure: 30 nmol of the sample was dissolved in 150 µl of 3 M MESA (4-morpolinoethanesulfonic acid, warmed up under hot tap water until it dissolved) and hydrolysed at 105°C, under vaccum as described above. Immediately after opening the glass tube, 700 µl of the resultant mixture and 350 µ l 1 M NaOH of internal standards were applied per column.

The amount of amin (nmols) acids were estimated on the basis of peak area by reference to standards of known amount. Residues per molecule

were calculated by using, as a divisor, the average figure of nmol per residue, giving close to integral values for the amino acids, and protein sizes estimated from SDS-PAGE. Only values of amino acids which are relatively stable or not forming bonds, particularly resistant to hydrolysis, were used. Thus, low Tyr, Pro, Ile and Val figures were excluded. The ratios obtained are considered to be more informative than the raw values. Residue fractions less than 0.2 residues/nmol were not considered significant. All figures above 0.1 residues/nmol are recorded.

2.3.8. The Generation of Peptides by Proteolysis

Digestion of cytochromes c was performed after dehaemation which ensured that they were fully denatured. This facilitates the action of proteases with the additional advantage of the haem site peptide being recovered in good yields. To obtain suitable peptides for sequencing the cytochromes, the following proteases were used: trypsin, which cuts on the C-terminal side of lysine; thermolysin, which cuts at the N-terminal side of hydrophobic amino acids, and the V8 protease, which cuts on the C terminal side of glutamate. The method used was the same in all cases: Apoprotein (approximately 1.5μ mol) was digested with 1/40 of its weight of protease, in 0.2 M ammonium acetate buffer, pH 8.5, for 4 hours at 37°C. Pepsin was used for analytical digestion to check for homology between different isolated cytochromes, and was performed on native proteins. 20 nmol of proteins were dissolved in 50 μ l of 5% formic acid and digested with 1/40 of its weight of pepsin, at 37°C for 4 hours. The digests were dried and analysed by HVPE, pH 6.5.

2.3.9. Purification of Peptides

Initial fractionation was done by gel filtration, followed by HVPE, pH 6.5. Sometimes, further purification steps were required, and were usually accomplished by successive HVPE at pH 3.5 or 2.0. In cases where HVPE was unsuccessful, paper chromatography was used. The paper was stained for detection of peptides. The techniques used can be found in Ambler (1963) and in Ambler and Wynn (1973). Before automated sequencing of some of the peptides was started, they were purified further by reverse phase HPLC.

2.3.9.1. Gel filtration

Gel filtrations were carried out on three Sephadex G-25 columns (1cmx90cm), arranged in series. The columns were equilibrated with 5% (v/v) formic acid. The freeze-dried digest was dissolved in 0.8 ml of buffer and applied on top of the first column. Flow rate of 20 ml/h was maintained by peristaltic pump, and the elutant monitored at 254 nm. At this wavelength, tryptophan absorbs strongest (e= 3000), with contributions from tyrosine (e=400) and phenylalanine (e=280). 1.5 ml fractions were collected and 150 μ l samples taken from each fraction for HVPE analysis on the distribution of peptides (see section 2.3.9.3).

2.3.9.2. High Voltage Paper Electrophoresis (HVPE)

These were carried out on a Whatman 3MM or 1MM paper (length= 57cm; width variable (cm) in Michl solvent-cooled tanks. Three different volatile systems, pH. 6.5, 3.5 and 2.0 were used (Ambler, 1963). Samples were spotted onto the paper and wetted by the solvent. Excess solvent was thoroughly mopped off. The peptides were run at 3 kV (potential gradient of 65 V/cm) for approximately 90 min and sometimes longer, depending on the separation required. At pH 6.5, the amino groups of the N-terminal end, and of lysine and arginine are positively charged and the carboxylic groups of the C-terminal end, and glutamates and aspartates are charged, so their mobility is determined by the net charge of the these groups. At higher pH, an increasing number of the carboxylic groups become protonated, histidine becomes positively charged, and the net charge of the polypeptides becomes more positive. At pH 3.5 the carboxylic groups of the glutamates and the aspartate have been neutralised, but the carboxylic group at the C terminal may still be negtively charged, but at pH 2 they are also protonated.

HVPE was used: to analyse the distribution of peptides in the gel filtration, for preparative separations of peptides following gel filtrations, for analysing the purity of peptides, and for comparative peptide mapping of different cytochromes c.

A marker mixture, "Wondermix", was always run in parallel to an unknown sample in HVPE pH 6.5 and pH 3.5. It contained 5 μ mol/ml Lys, His, Arg, Gly, Val, Glu, Asp, alanyl glycine, taurine, cysteic acid, DNP lysine (yellow) and xylene cyanol FF (blue dye). For HVPE pH 2, the markers T and R were used.

2.3.9.2.1. pH 6.5 electrophoresis

Solvent is pyridine/acetic acid/water, in the ratio 25:1:225, by volume. The coolant is toluene. Samples were spotted on to Whatman 3MM paper, 9 inches from the end submerged into the cathode buffer of the system. Yellow orange was used as a marker, to monitor the progress of the electrophoresis run. A mixture of standard amino acids with characteristic mobilities, so called Wondermix, was run in parallel.

2.3.9.2.2. pH 3.5 electrophoresis

Solvent is pyridine/acetic acid/water, in the ratio 1:10:225 by volume, pH 3.5. The coolant is white spirit. Samples were spotted 3.5 inches from the end submerged into the cathode buffer. Methyl green was used to monitor the progress of the run, and the R- and T-mixtures were used for mobility comparisons to peptides. At this pH, the sidechain carboxylic groups of glutamate and apartate become protonated and therefore neutral, but the amino groups positively charged.

2.3.9.2.3. pH 2.0 electrophoresis

The solvent was formic acid/acetic acid/water, pH 2.0, in the ratio 1:4:45 by volume. White spirit was used as the coolant. Samples were spotted on a 3MM Whatman paper (or 1MM Whatman for qualitative amino analysis)12 cm from the end submerged in the cathode buffer. Methyl green and Wondermix were used as markers in peptide separations T and R marker mixture and methyl green was used for qualitaitive amino acid analysis.

2.3.9.3. Mapping of peptide distribution, in gel filtration elutates by HVPE

A 150 µl sample was taken from each gel filtration fraction, for HVPE analysis. They were dried down and dissolved in 20 µ l of 0.1m 1N NH_3. Then, 10 µl at time, were spotted in the order of elution fractions, on to a Whatman 3MM paper, and separated in parallel by HVPE, at pH 6.5. The paper was stained for the peptides and the resultant map gave the elution profile of the peptides.

2.3.9.4. Preparative separation of peptides by HVPE

After gel filtration, HVPE was used as the final purification step. The peptide map was used to decide which fractions to pool. The intensity of

the peptide spots was used for rough estimation of suitable peptide concentrations for the preparative electrophoresis. Pooled fractions were freeze-dried, and dissolved in 0.1 M NH_3 to give approximately 2 μ mol/ml. The solutions were then loaded onto a paper at 10-100 nmol/cm. In most instances, electrophoresis in one solvent system was sufficient to obtain a pure peptide, but sometimes two consecutive runs in different systems (pH 3.5 and/or pH 2.0) were necessary. Marker strips, containing a portion of the solution, and a standard mixture (Wondermix), were run on either side of the main peptide area, which was cut out and stained. These were used to delimit the separated peptide bands which then were cut out for the subsequent elution. At this stage, peptides were, while still in the paper, sometimes oxidised. The peptide bands were cut out and eluted with 0.1 M NH_3 into test tubes in specific elution chambers. Very basic peptides were eluted with 1 M acetic acid. Purity of the peptides was assessed, by running samples in two different solvent systems, pH 3.5 and pH 6.5. If a single spot emerged in both runs after staining, the peptide was assumed to be pure. This was further substantiated by a N-terminal analysis and quantitative amino analysis.

Common contamination of peptides was by the amino acids glycine and serine and, to a lesser extent, of small, neutral peptides with comparably charged peptides. In most cases, these contaminating peptides were in such negligible quantities so as not to interfere with the sequencing results.

2.3.9.5. Paper chromatography

The method of descending paper chromatography was used to separate small neutral peptides after thermolysin digest. The composition of the solvent was: butane-1-ol/acetic acid/pyridine/water in the ratio 15:3:10:12 by volume. The duration of the chromatography was 12 hours at 20°C. Peptides were loaded in 2 μmol/ml, 50 nmol/cm onto a Whatman 3MM paper.

2.3.9.6. Methods for detecting and characterising peptides and amino acids on paper

Fluorescence under UV light

Following HVPE separation, the peptides were always viewed under 340 nm UV light, as tryptophan-containing peptides are weakly fluorescent under this wavelength. This fluorescence is enhanced if the peptides are

oxidised by performic acid. Tryptophan is destroyed by this treatment to yield highly fluorescent products.

Staining with pyridine/collidine for detection of peptides and free amino acids

Pyridine/collidine staining was used to detect peptides and free amino acids on paper. Ninhydrin reacts with free amino- (or imino) group. Collidine interacts with this compound, to give distinct characteristic colours to free amino-acids which change, with time to blue. The mixture gives yellow colour to peptides with N-terminal glycine, threonine, aspargine, serine, methionine sulphoxide, cysteic acid and proline. All but the proline spots, slowly change to blue. Other peptides usually develop blue colour but sometimes they also become yellow. The sensitivity is approximately 5 nmol/cm^2, but is dependent on the number of amino groups in the peptide. These are, in addition to the α-amino group of amino acids and peptides, the ε-amino groups of lysine and arginine.

Procedure: Following HVPE, the paper was allowed to dry. It was then immersed in a mixture of 0.1% ninhydrin in acetone and 1% (peptide maps, marker strips) or 4% (qualitative amino acid analysis) v/v 2,4,6 collidine in acetic acid (1:2 v/v). The paper was allowed to dry at room temperature and then heated at either 60°C or 105°C, until the colours started to develop. Care was taken not to heat it for too long as the colours indicating the presence of particular amino acids, will change with time and this is accelerated by the heat.

The Pauly test

This method was used to detect peptides which contain histidine, directly after the ninhydrin/collidine staining.

The paper was sprayed with a mixture of 1% sulphanilic acid in 1 M HCl v/v and 5% aqueous $NaNO_2$ (W/v) in the ratio 1:1 by volume, which had been cooled at 4°C for 15 min. The paper was then sprayed with 15% Na_2CO_3 (w/v). Peptides, containing histidine, slowly developed rose red colour while other peptides became orange. The sensitivity was about 5 nmol/cm^2. Histidine may indicate which peptide contains the haem site.

2.3.10. Reverse Phase HPLC

Peptides, purified by HVPE, were usually purified further by reverse phase high pressure liquid chromatography (HPLC), before automated sequencing was started. The equipment used in this work was from Water Associates (Instruments) and comprised two model 510 pumps, a model 680 automated gradient controller, a model Lambda-Max absorbency detector, and a model U6K universal injector. The mBondpack C_{18} column (10 mm particle size; 0.39 cm x 30.0 cm) was from Millipore. All samples for the HPLC were centrifuged (10000 g for 5 min) in a microcentrifuge to remove particulate material. The gradient controller was programmed to include a 5 min 'loading time' during which the peptide and protein samples were applied isocratically (i.e. 100% solvent A) and the flow rate was increased linearly from 0.1 to 1.0 ml/min. The peptide peaks were resolved by using linear gradient, starting with 100% solvent A , 0-40% solvent B, for 60 min and 40% -60% solvent B for 30 min. Solvent A: 0.1% trifluoroacetic acid (TFA). Solvent B: 0.1% TFA in acetonitrile (Grade S, Rathburn, U.K.).

2.3.11. Dansyl-Edman Sequencing of Polypeptides

The Edman degradion method was used to remove amino acid residues from peptides and proteins, one at a time. Each removal is defined as one cycle of degradion. Sample was taken of the original peptide and after each cycle, and the N-terminal residue identified by the Dansyl-chloride method.

2.3.11.1. Dansylation

Procedure: 5-10 nmol of peptide in water were dried down in 3.5mm x 28mm glass tube under vacuum. 10 µl of 0.1% (w/v) sodium bicarbonate was added and immediately dried down. 10 µl of water and 10 µl of 0.5% (w/v) dansyl chloride in acetone were added, the tube covered with parafilm and incubated at 37°C for 30-60 min. This mixture was then dried down under vacuum, and 50 µl of 6 M HCl added. The tube was sealed and incubated for 16-24 hours at 105°C. The incubation time was reduced to four hours if proline was expected to be at the N-terminal. The tube was then opened, the hydrolysate dried down under vacuum, and redissolved in 10 µl of 50% (v/v) pyridine. The dansylated terminal residue was

identified by HVPE pH 4.0 by comparison to standard mixtures of dansyl-chloride amino acids, which were run at the same time.

2.3.11.2. Identification of dansyl-chloride amino acids by HVPE pH 4.0

The electrophoresis was carried out in a specific apparatus, where the paper is pressed between two water cooled plates. The solvent buffer was pyridine/acetic acid/H_2O, 10:1:89 by volume. The paper sheet was approximetely 70 cm long and the samples were spotted onto the paper, 33 cm from the end which was submerged into the cathode buffer. The paper was wetted in the solvent buffer, and excess buffer thoroughly mopped off. The samples were run for approximately 3 hours. Dansylated amino acids, which were divided into two standard mixtures, P and Q, were run in parallel to enable the identification of the amino acids in the unknown sample. The P-mixture contained DNS-cysteic acid, DNS-glutamic acid, DNS-OH (blue), DNS-serine, DNS-alanine, DNS-leucine, DNS-isoleucine, DNS-a-tyrosine, DNS-a-histidine (cathodic), DNS-arginine (cathodic) and DNS-NH_2 (cathodic). The Q-mixture contained DNS-aspartic acid, DNS-OH (blue), DNS-proline, DNS-threonine, DNS-valine, DNS-phenylalanine, bis(a,E) DNS-lysine (smear at origin), and DNS-NH_2. The order given, is the order of separation from the anode (+) to cathode (-) on electrophoresis at pH 4.

2.3.12. Vapour-Phase Pyridylethylation of Cysteine Residues

The peptide was spotted on to a glass-fibre filter disc and dried under a stream of N_2. The pyridylethylation reaction was done in a stoppered glass tube (50mm x 10mm) with a constriction in the middle, to support the filter. Pyridine (100 µ l), water (100 µl), 4-vinylpyridine (20 µl) and tributylphosphine (20 µ l) were placed in the lower part of the tube. The tube was flushed thoroughly with argon, and the peptide-coated disc placed on the constriction before the tube was sealed with a stopper. The vapour-phase reaction was allowed to continue for two hours at 60°C. Subsequently, the disc was removed from the tube, avoiding touching any liquid, and washed for 10s in 2 ml of each of the three following solvents: n-heptane, n-heptane/ethyl acetate (1:1 v/v) and ethyl acetate. The washed disc was allowed to dry at room temperature before wetting with 30 µl (2 mg) of Polybrene. The Polybrene treated disc was redried and placed in the cartridge block of the 477A sequencer. To remove all traces of

reagent, the sample was precycled with a wash program (n-heptane, followed by butyl chloridem 12.5% trimethylamine in water, n-heptane and ethyl acetate) before the normal sequencing programme was begun. If the disc is contaminated with pyridine it is manifested as a broad peak in the middle of the chromatogram from PTH C_{18} column. The pyridylethylcysteine- phenylthiohydantoin derivative is resolved on the PTH C_{18} column, with the standard separation protocol used by the on-line model 120 phenylthiohydantoin analyser, and is eluted just before diphenylthiourea.

2.3.13. Automated Protein Sequencing

Peptide fractions from the HVPE peptide purification were examined for purity before being sequenced, using an Applied Biosystems 130A Microbore Separation System. The column used was an Aquapore RP-300 (7 mm particle size; 2.1mm x 30mm), also from Applied Biosystems, which was developed with a linear gradient, 8-80% (v/v) of acetonitrile in aq. 0.1% (v/v) trifluoroacetic acid, and monitored at 220 nm. The pure peptides were subjected to automated sequencing on an Applied Biosystems 477A instrument with a 120A on-line phenylthohydantoin analyser. Polybrene (2mg) was loaded on to a glass-fibre filter disc, and pre-cycled three times before being loaded with approximately 1.5nmol of the peptide. The peptide was pyridylethylated before application. After Edman degradation, the anilinothiazolinone derivatives, cleaved from the peptide, were converted automatically into the more stable phenylthiohydantoin forms and separated on an Applied Biosystems PTH C_{18} column (5 mm particle size; 2.1mm x 220mm) which was developed with a 0-100% (v/v) gradient of acetonitrile. The gradient was formed by using an aq. 5% (v/v) solution of tetrahydrofurane as solvent A, and acetonitrile as solvent B. Chromatography was performed at 55°C and the column monitored at 269 nm.

2.4. DNA TECHNIQUES

Unless specifically referred in the appropriate sections, the basic DNA protocolls that were used are found in the following standard protocol

manuals: Molecular cloning. A laboratory manual, 1982 and 1989, Cold Spring Laboratory Press, USA; The guide to molecular cloning techniques. Methods in Enzymology. Volume 52. 1987, Academic Press, San Diego ; DNA cloning: A practical approach. Volumes I-III. IRL Press at Oxford University; Nucleic acid hybridisation: A practical approach. 1985, IRL Press at Oxford University and in Nucleic acid sequencing. A practical approach. 1989, IRL Press at Oxford University. Enzymatic reactions were carried out according to the protocols supplied from the manufacturer of any particular enzyme.

2.4.1. Preparation of Genomic DNA From *Bacillus licheniformis.*

Cells were grown in L-broth overnight. They were harvested by centrifugation at 12000 g for 5 min. and washed twice in TE buffer (10 mM Tris/HCl pH 8, 1 mM EDTA). The packed cells (3-6 g) were resuspended in 6 ml of saline-EDTA (0.15 M NaCl, 0.1 M EDTA, pH 8.0) and 12 mg lysozyme added, as well as proteinase K (0.01mg/ml) and heat-treated RNAase (0.01 mg/ml). The mixture was incubated at 37°C for 20-30 min. When the cells were just beginning to lyse, the mixture was quickly frozen. Then 50 ml of Tris-SDS buffer (0.1 M Tris/HCl, pH 9.0, 0.1 M NaCl, 1% SDS) was added. Thawing and freezing were continued until complete lysis had occurred. Then, the mixture was mixed with an equal volume of phenol, saturated with Tris-SDS buffer (pH 8.5, 1% SDS) and carefully shaken in the cold for 20 min. At the end of this time, the suspension was given a quick very thorough shake. If this was not done, the DNA would become mixed with a transparent sticky substance which would interfere with successive purification steps.

The emulsion obtained was separated into two layers by centrifugation, at 4000 g for 5 min in GSA rotor. The upper aqueous phase was clarified by centrifugation at 15000 g for 15 min and extracted once with one volume of an equal volume mixture of phenol, saturated with Tris-SDS buffer and chloroform/isamyl alcohol (24:1), and once with one volume of chloroform/isoamyl alcohol. The DNA was precipitated with two volumes of 99% ethanol, after adding 0.05 volumes of 3 M sodium acetate, pH 4.5. The pellet was rinsed with 70% ethanol. The DNA was dissolved in TE buffer to give concentration of approximately 1 μ g/μl.

2.4.2 Size Selection of Genomic DNA

Procedure: 150 µg of genomic DNA was digested with an appropriate restriction enzyme and loaded into a 4 cm slot on a 14 cm long horizontal agarose gel, and run overnight at constant 25 Volts. A gel slice with DNA, corresponding to 1 Kb size range region and including the gene containing fragment, was cut out and the DNA extracted by the Geneclean method. An alternative method was used as well, the selected DNA size range region was then run into a DEAE paper (section2.4.7.2).

2.4.3. Determination of DNA Concentration and Purity

The absorbency of DNA, at 280 and 260 nm was measured. The concentration, µ g/ml, of double stranded DNA, in the sample is given by the relation A260x50. The concentration of single stranded oligonucleotides is given by the relation A260x20. The ratio of 260:280 gives the measure of purity of the DNA. DNA with values >1.7 were considered sufficiently pure and therefore used.

2.4.4. Restriction Analyses of DNA

DNA from *B. licheniformis* was restricted with endonucleases which recognise a specific six-base sequences and which, upon cleaving, give fragments with staggered ends. 0.5-5.0 µ g of DNA was cut in a final volume of 10-20 µl. Attempts were made to optimise conditions by using prescribed buffers for every enzyme, and spermidine (5 mM) and BSA (0.01%) in the reaction mixtures. The DNA was also diluted in the reaction mixtures (to 0.25 µg/µl) to minimise the danger of possible contaminants in the DNA, which might affect the efficiency of the restriction enzymes. The possible thermolability of the enzymes was taken into account by adding another portion of the enzymes, after 1-4 hours had passed. Only two of the enzymes cut the DNA satisfactorily: *Hind*III and *Cla*I. Consequently, this restricted the scope of further experiments. In subsequent restriction reactions the One PhorAll buffer from Pharmacia was used. Restriction analysis of various cloning vectors was performed under same conditions, but using less concentration (from 1-5 units) and for a shorter time (2-4) hours.

2.4.5. DNA Electrophoresis on Agarose Gels

Analysis of restricted DNA was carried out on a horizontal 0.7% (w:v) agarose 1x TBE (0.045 M Tris/borate 0.001 M EDTA) or 1x TAE(0.04 M Tris/acetate, 0.001 M EDTA) gels, submerged in the same buffer. The fluorescent dye, ethidium bromide was added (0.01 mg/ml). Minigels of the dimensions 70mm x 9mm x 3mm were used for quick analyses of restriction efficiency of restrictions, and to estimate the sizes of inserts in recombinant plasmids. Loadings were approximately 1 μ g of DNA. Bigger gels of the dimensions 150mm x 90 mm x 10mm were used, to analyse the fragment distribution in genomic restrictions, for Southern blottings, for isolations of specific size range of genomic DNA fragments and, the isolation of insert DNA of recombinant clones. Loadings were up to 150 μg/well, with a 30 μ g maximum per cm.

DNA was mixed with 1/3 volume of loading buffer (40% sucrose, 0.25% w:v bromophenol-blue 7x TBE). The minigels were run at a constant 50-100 voltage for 2-4 hours. The larger gels were run at a constant voltage of 25 overnight, or until the bromophenol-blue had migrated to the bottom of the gel. Gels were viewed under UV light and, when necessary, photographed.

2.4.6. Southern Blotting of Restricted Genomic DNA

The DNA in the agarose gel was denaturated by soaking the gel for one hour in several volumes of 1.5 M NaCl, 0.5 M NaOH, with constant gentle agitation. It was rinsed briefly in distilled water, and then neutralised twice, for 30 min and 15 min, in several volumes of 1 M Tris (pH 7.4), 1.5 M NaCl, at room temperature It was then placed on to a flat plexiglass platform which had been wrapped in Whatman 3MM paper and prewetted in transfer buffer. The platform was placed into a shallow tray which was filled with the transfer buffer, 20x SSC (175.3 g/l NaCl, 88.2 g/l Na_3Citrate, pH 7.0) until the level of the liquid almost reached the top of the support. A nitrocellulose or a H-bond nylon filter which had been cut to the dimensions of the gel, prewetted in distilled H_20 and immersed in the transfer buffer for 5 min, was placed on top of the gel. Onto the nitrocellulose filter were placed two sheets of Whatman 3MM paper, cut to the same dimensions and then a stack of paper towels, 5-8 cm high. A glassplate and a 500 g weight was put on top of the stack. After 16 hours

the gel filter was peeled off and soaked in 6xSSC for 5 min at room temperature, then dried for at least 30 min at room temperature. To fix the DNA to the filters the nitrocellulose filters were baked for 2 hours at 80°C under vacuum, but H-bond nylon filters were exposed to UV for 1-5 min.

2.4.7. Recovery of DNA from Agarose Gels

Restricted DNA was run as described in section 2.4.5, in low EEO agarose gels, as this type of agarose gives sharper bands of fragments, ensuring separation of fragments of similar size. Two methods were used for extracting DNA from the gels (sections 2.4.7.1 and 2.4.7.2.).

2.4.7.1. The Geneclean method

The Geneclean kit from Bio 101 Inc. was used. This kit contains a specifically formulated silica matrix which binds single and double stranded DNA. Up to 80% recovery was obtained. The protocol of the manufacturer was used.

2.4.7.2. The DEAE-filter method

DNA (10-30 μg/cm) was run in TBE gels until clearly resolved, as visualised under UV light. The gel was then cut above the target DNA and removed. Another transverse cut was made directly underneath the target DNA and to one side of the gel. Into the cut was placed a DEAE-filter strip, prewetted in NET buffer (150mM NaCl, 0.1mM EDTA 20mM Tris-HCl, pH 8.0) of the same width as the DNA but 1 mm higher than the gel. The electrophoresis was then continued for 10-30 min or until the DNA had run into the filter. Then the filter was removed, cut into 1 cm bits and two bits put into one or more Eppendorf tubes. 300μ l of HNET (6M NaCl, 0.1mM EDTA 20mM Tris-HCl, pH 8.0) buffer was added to each tube and incubated for 20 min. Every 5 min the tubes were vortexed. The liquid was then pooled into one Eppendorf tube, extracted first with phenol/chloroform and then with chloroform only. The DNA was ethanol precipitated and washed with 70% ethanol (K. Lineruth personal communication)

2.4.8. Plating of Lambda-Phages

100 µl of an appropriate dilution of phage solution, was used to infect 200 µl of an overnight culture of *E. coli* NM 514 (see section 2.1.2), to which 30 µl of 100 mM $MgSO_4$ had been added. The culture was left at room temperature for 15-20 min, then mixed with 3 ml of BBL-top agar and left to cool to about 45°C. The mixture was poured on a BBL-plate, which was inverted and incubated for 37°C overnight.

2.4.9. Large Scale Preparation of Lambda-DNA

2.4.9.1. Plate lysate

A single plaque (approximately 10^6 pfu) was transferred by Pasteur pipette into 1 ml phage buffer ($KH2PO_4$ 3 g/l; Na_2HPO_4 (anhydrous), 7 g/l; NaCl, 5 g/l; 10 ml/l 0.1 M $MgSO_4$; 10 ml/l 0.01 M $CaCl_2$ 1 ml/l 1% gelatin, pH, 7.2). A drop of chloroform was added and the suspension vortexed briefly. 0.1 ml of this solution was added to 0.2 ml of appropriate *E.coli* strain, and incubated for 10 min.at room temperature to ensure absorption. 3 ml of BBL-top agar was added, and the mixture poured quickly onto the surface of an 82 mm L-broth plate and left to solidify. The plate was incubated for eight hours at 37°C without inverting, when confluent lysis has occurred. Then the loose agar top was scraped off with a sterile spatula into a centrifuge tube, and vortexed with several drops of chloroform. The lysate was centrifuged at 8000 g for 10 minutes and the supernatant saved and titrated.

2.4.9.2. Liquid lysate

An overnight culture of *E.coli* cells were diluted to 1:40 by $MgSO_4$ supplemented (0.01 M) L-broth, to give 200 ml, and grown at 37°C with vigorous shaking to ABS 650 nm of 0.5 OD. This OD corresponds to concentration of approximately $2.5x10^8$ cells/ml. To this culture was added a sufficient volume of phage plate lysate to give a 0.1 ratio of phage particles/cells unit volume and incubated for further four hours at 37°C. During this time, the optical density of the culture reached approximately 2.0 OD units, before falling sharply to 0.5 OD, which indicates that phage induced lysis had taken place. Complete lysis was ensured by shaking the suspension with 0.5 ml of $CHCl_3$. Cell debris was removed by centrifugation

for 15 min at 6500 g (GSA-rotor) and at this stage the lysate was usually titrated. Purification was continued if the titre had reached 10^{11}.

To the lysate (200ml), 8 g of NaCl and 28 g of PEG 6000 were gradually stirred in until it had dissolved at room temperature. This solution was incubated overnight at 4°C and then centrifuged at 5000 rpm (GSA-rotor) for 20 min. The supernatant was discarded and the pellet resuspended in 5 ml of phage buffer by shaking, at 4°C for two hours. The suspension was centrifuged for 5 min at 2500 g (GSA-rotor) and the supernatant layered onto a step gradient (1.5 ml of 1.7 g/ml-2 ml of 1.5 g/ml-2 ml of 1.3 g/ml, CsCl) in a tube for a 6x14 ml rotor (Du Pont) and centrifuged for two hours at 100000 g and 20°C. The phage band was collected with a 21 gauge syringe by a side puncture. The phage solution was dialysed against TE buffer for one hour, phenol extracted three times and back-extracted with and equal volume mixture of phenol and chloroform, and with chloroform only.

2.4.10. Plating of Bacteriophage M13

Bacteriophage M13 was plated on lawns of *E. coli* NM522 or TG1 cells. 100 µl of the appropriate dilution of phage solution was mixed with 200 µ l of fresh overnight culture of plating cells and 3 ml of BBL top agar molten, and cooled down to about 45°C, then plated on L-plates. The plates were incubated overnight at 37°C. When a blue to white selection of recombinant phages was required, 30 µ l/3ml top agar of isopropyl-β-D-galactoside (IPTG) and 30 µl of the X-gal were used (see section 2.1.3).

2.4.11. Isolation of DNA From M13 Phages

2.4.11.1. Isolation of single stranded DNA

Single plaques of M13 mp19 were transferred to 100 ml culture of log phase TG1 cells (OD at ABS 650 nm, 0.3) and incubated with shaking for five hours at 37°C. 1.5 ml of the culture was then transferred to an Eppendorf tubes and spun down in microcentrifuge. The supernatant was decanted to a fresh Eppendorf tube and the cell pellet set aside for the isolation of a double stranded DNA (see next section). 150 µl of 20% Polyethylene glycol (PEG 6000) 2.5 M NaCl were added to the supernatant and incubated at 22°C for 10 min. The precipitated phage was collected by centrifugation for 10

min in a mirocentrifuge. The supernatant was poured off and the tube spun again for 5 min. The last drops of the Peg solution supernatant were removed by careful suction, using a drawn out Pasteur pipette. The pellet was resuspended in 100 µl of TE buffer (10 mM Tris-HCl, pH 7.61, 1 mM Na_2EDTA) and the suspension extracted with 100 µ l of neutralised phenol solution. The suspension was reextracted with 100 µl of an equal volume mixture of neutralised phenol and chloroform/isamyl alcohol (24:1). The DNA was precipitated with two volumes of 99% ethanol, after adding 0.05 volumes of 3 M sodium acetate, pH 4.5. The pellet was rinsed with 70% ethanol and dissolved in 20 µ l of TE buffer which gave enough template for 3-4 sequencing reactions.

2.4.11.2. Isolation of double stranded DNA

The cell pellet from the single stranded DNA isolation above (section 2.14.11.1) was resuspended in 100 µl lysis solution (50 mM glucose, 20 mM Tris-HCl, pH 8, 10 mM EDTA, 0.1% w/v lysozyme, added prior to use) and incubated for 5 min at room temperature. 200 µ l of freshly prepared SDS solution (0.2 M NaOH, 1% w/v SDS) were added and the mixture left on ice for 5 min. 150 µl of preecooled 3 M potassium acetate, pH 4.8, were then added and the resulting mixture put on ice for 5 min. The precipitated chromosomal DNA, proteins and SDS were removed by centrifugation for 10 min and the supernatant transferred to a clean tube. This step was repeated to ensure that no precipitate was carried over to the next step. The nucleic acid in the supernatant was precipitated with two volumes of ethanol, and washed with 70% ethanol, vacuum dried and resuspended in 20 µl of TE buffer. Sometimes, an extra step was included before ethanol precipitation. Then the mixture was extracted with an equal volume mixture of phenol/chloroform to inactivate DNAases

2.4.11.3. Large scale preparation of single stranded M13 sequencing template

This method is essentially a scaled up version of the foregoing method with minor modifications. It yields 50-100 µ g of high quality single stranded DNA.

A phage plaque was transferred into 2 ml of a log phase *E. coli* TGH 1 cell culture and grown overnight at 37°C. The 2 ml culture was transferred to 100 ml of log phase *E.coli* TG1 cells and incubated with shaking for six hours at 37°C. After removing the cells at 300x g for 5 min, the bacteriophage was precipitated by adding 20 ml of 20% polyethylene

glycol (PEG 6000), 2.5 M NaCl solution. and incubated overnight at 4°C. The precipitated phage was collected by centrifugation and resuspended in 1 ml of TE buffer. The suspension was transferred to an Eppendorf tube and remaining cells removed by centrifugation for 15 min in a microcentrifuge. The phage was precipitated again by adding 20 µ l of the PEG solution, incubated for 10 min at 22°C and collected by centrifugation for 10 min in a microcentrifuge. The procedure from then on is the same as for the small scale preparation. The resulting single stranded DNA was dissolved in TE buffer, giving a concentration of approximately 1 µ g/µl.

2.4.12. Blotting of Libraries onto Nitro-Cellulose and H-Bond Nylon Filters

Plated out phage-libraries with densities up to 10 plaques/cm^2 were either used directly, or plaques were transferred to a second plate onto a grid. The plates were placed at 4°C for at least one hour. Nitrocellulose filters were prewetted in 4x SSC and H-bond nylon filters were prewetted by putting them onto a L-agar plate. This was necessary to prevent the top agar from sticking to the subsequent plaque lifting. The filters were put carefully onto the phage library plates for 1 min and duplicate filters for 2 min. The filters were then transferred onto a blotting paper, soaked in 20x SSC for 5 min and then floated onto a 0.5 M NaNOH 1.5 M NaCl solution in a shallow tray. They were then laid onto a blotting paper, soaked in 0.5 M Tris/HCl, pH 7.5, 1.5 M NaCl, for 3 min. Nitrocellulose and nylon filters were dried for 30 min and the DNA fixed by baking the nitrocellulose filters and exposing the nylon filters to UV light for 2-5 min.

2.4.13. Radiolabelling of Oligonucleotide Probes

Labelling reactions were done in 10 µ l volume, containing: 2 µ l (20mCi) g-(^{32}P)ATP (3000Ci/mol); 3 µl 1 M Tris/HCl, pH 8; 1.5 µl 0.1 M DTT; 1 µl 0.3 M $MgCl_2$; 20 pmol oligonucleotide H_2O to 30 µl; 0.5 µl (5 units) T4-polynucleotide kinase. The mixture was incuabted at 37°C for 45 min. The enzyme was inactivated by heating the mixture for 10 min at 68°C. The efficiency of the kinasing reaction was checked by the CTAB method (P. Ford, personal communication). The kinasing reaction was diluted 100 fold and 1 µl mixed with 100 µl of H2O. 0.2 ml of yeast RNA solution (2 mg/ml

yeast RNA; 1 M Na acetate/acetic acid, pH 4.8) and 1 ml 4% CTAB (cetyl trimethyl ammonium bromide in distilled H_2O). A fine precipitate formed on standing for 5-10 min at room temperature. The precipitate was collected by water suction onto glass filter disc and washed extensively with H_2O. The disc was dried under vacuum and the radioactivity was counted by scintillation. Typically, counts of $2x10^8$ cpm/ml labelling mixture, were obtained, which is adequate.

2.4.14. Hybridisation of ^{32}P-Labelled Oligonucleotide Probes to DNA on Nitrocellulose and H-Bond Nylon Filters

To reduce background, the filters were prehybridized at 60°C for 2-16 hours in 4x NET, 0.01% Na-pyrophosphate, 0.1% SDS, 0.01 M EDTA. Hybridization was carried out in the same solution with 1-50 ng/ml of radiolabelled probe. Filters were washed at room temperature with 6x SSC and then at increasing temperatures until a single or a minimum of bands or plaques stood out from the lighter background. The washing was continued until no further difference in intensity could be achieved between bands or plaques, in some cases with genomic blots, just short of washing everything off.

Washing with tetramethylammonium chloride (TMAC) results in better discrimination between bands and plaques due to the elimination of the preferential melting of AT versus GC base pairs with SSC. The stringency of the hybridization is then a function of probe length only (Wood *et al.*, 1985). The procedure was the same as for washing 6 SSC, and 3 M TMAC was used. This was the preferred washing method in the latter stages of the project.

2.4.15. Dephosphorylation

To reduce screening background when DNA fragments were cloned into M13 derived vectors, self-ligation was prevented by removing the 5'-end phosphates of the restricted vectors. The following procedure was used: A restricted cloning vector with 10 pmol of ends, was mixed with 0.1U calf intestine phosphatase (CIP) in a 50 µ l volume of CIP buffer (500 mM Tris/HCl, pH 9; 10 mM $MgCl_2$; 1m $ZnCl_2$; 10 mM spermidine). The mixture was incubated for 30 min, another 0.1 U of CIP was added and the

incubation continued for additional 30 min. The reaction was stopped and the enzyme destroyed by incubation at 80°C for 15 min.

2.4.16. DNA Ligation

The restricted DNA was diluted to about 100 ng/µl. The reaction was carried out in 10 µl volume containing the vector and the insert DNA, 1 µ l 10 mM ATP, 1 µl 100 mM DTT, 1 µOnePhorAll buffer (Pharmacia) pH 7.5, 1 µl 100 mM MgCl2, 1 µl 2 mM spermidine and T4 ligase. Typically 0.01 units of ligase were used for sticky ended ligation in a 10 µ l reaction mix. The ligation was incubated overnight at 15°C. A series of controls, with the DNA components in different ratios and without either the insert or the vector, were carried out. Half of each reaction mixture was used to transform an appropriate *E. coli* strain. One control also included dephosphorylated vector DNA to check the efficiency of the preceding restriction step. If the restriction was efficient, no ligation would occur and no transform, the host would be observed.

2.4.17. The Construction of Genomic Library

2.4.17.1. Transformation with λ-DNA

2.4.17.1.1. Preparation of competent cells for transformation with recombinant λ- phage. A fresh overnight culture of *E. coli* 514NM was used to inoculate 200 ml L-broth, (1:50 v/v) containing 20 mM $Mg/Cl_2/SO_4$ (an equimolar mixture), and incubated at 37°C with vigorous shaking until the OD_{540} was between 0.5-0.6. The culture was chilled on ice, transferred to a sterile centrifuge bottle, and the cells pelleted at 2500 g for 10 min at 4°C in a GSA rotor. The supernatant was discarded and the pellet resuspended in the 1/4 of the original culture volume of ice-cold 0.1 M $MgCl_2$. The cells were pelleted again, the supernatant poured off, and the cells resuspended in 1/20 volume of the original volume of ice-cold $MgCl_2$. The cell suspension was transferred to a sterile 25 ml Universal bottle and kept on ice for 45 min. The cells were re-pelleted by a micro-centrifuge at full speed for 10 min at 4°C, the supernatant discarded and the pellet suspended in the same volume of 0.1 M MOPS (3-(N-

morpholino)propanesulfonic acid), pH 6.5, 50 mM $CaCl_2$, 20% glycerol. The suspension was kept on ice for 10 min, divided into aliquots and frozen.

2.4.17.1.2. Transformation. 100 µ l of thawed competent cells were mixed with 5 µl of recombinat λ-vector DNA and 100 µ l of SSC/$CaCl_2$ (4 ml 0.1 M $CaCl_2$/3 ml 1xSSC, and the mixture incubated for 5 min at 37°C. 3 ml of BBL-top-agar were added and the mixture poured on top of moderately humid L-plates. The recombinant phages were recognised by their clear plaque formation.

2.4.17.2. The λ-DNA *in vitro* packaging method

A λ1149-phage library was constructed from the *B. licheniformis* genomic DNA, using the λ- DNA *in vitro* kit from Amersham. In vitro packaging of recombinant λ- vectors, involves the use of cell extracts from two induced λ-lysogens with prophages that carry different, but complementing mutations in the genes, required for the assembly of mature phage particles. These extracts were mixed together with DNA of appropriate size and form, and the DNA packed into infectious phage particles. These phages were then introduced into the host cells by the normal process of infection. The library was constructed,according to the prescribed protocol from the manufacturer of the kit. The host *E.coli* NM514 was used for this (see section 2.1.2).

2.4.18. DNA Sequening

2.4.18.1. Cloning into sequencing vectors

The M13mp18, and M13mp19 vectors were used as sequencing vectors. The same restriction enzyme was used to cut the DNA insert the recombinant λ 1149-phage (1 µg/ml) and to restrict the dephosphorylated sequencing vector (1 µ g/ml). After inactivation of the restriction enzyme, usually by heating at 80°C for 15 min, 1 µ l of each restriction mixture was mixed together for ligation. The *E. coli* 522NM* was transformed with the ligation mixture. The transformation mixture was plated out, in 3 ml BBL-top agar onto Luria-broth, and incubated at 37°C overnight. As the host is resistant to λ-phages and the sequencing vector dephosphorylated, only recombinant M13 phages became visible as clear plaques.

2.4.18.2. Extended DNA sequencing

The extended sequencing method (Roe *et al.* 1985), which is a modification of the standard Sanger nucleotide sequencing method, was used for sequencing. This method made it possible to read in excess of 750 nucleotides from the priming site. In this method, the chain extension reaction is performed separately from the dideoxynucleotide termination reaction.

Sequencing buffer: (5 x concentrate), 200 mM Tris/HCl, pH 7.5; 50 mM, MgCl2; 250 mM NaCl.

Extension mix: (5 x concentrate) 7.5 mM dCTP, 7.5 mM dGTP, 7.5 mM dTTP (prior to use, diluted to 1:5 with H_2O).

Termination mixes:

A: 300mM ddATP, 25mM dCTP, 250mM dCTP, 250mM dGTP, 250mM dTTP.
C: 100mM ddCTP, 25mM dATP, 250mM dATP, 250mM dGTP, 250mM dTTP.
G: 150mM ddGTP, 25mM dGTP, 250mM dATP, 250mM dCTP, 250mM dTTP.
T: 500mM ddTTP, 25mM dTTP, 250mM dATP, 250mM dGCTP, 250mM dGTP.

Annealing reaction.

1 µg (0.4 pmol) of template was incubated with the universal sequencing primer, 0.4 pmol, in a total volume of 10 µ l sequencing buffer, in a 1.5 ml Eppendorf tube, at 60°C for 5 min. The annealing mix was then cooled down to room temperature (30 min) and concentrated to the bottom of the tube by centrifugation.

Chain extension reaction.

To the 10 µl of the annealing mixture were added 2 µ l of 5x extension mix, 1 µl of ^{35}S-dATP, 0.5µl of Klenow polymerase, large subunit, 5 u/µ l and 3.5 µl distilled water. The mixture was incubated for 10-15 min at room temperature.

Termination reaction.

After the chain reaction, 3.5 µ l of the extension mix were added to each of four Eppendorf tubes, labelled T, C, G, and A, with 2 µ l of corresponding termination mix. The tubes were incubated for 5 min at 31°C. The reaction was stopped by adding 4 µl of formamide dye (30 mg xylene cyanole, 30 mg bromophenol-blue, 0.2 ml of 0.5M EDTA, 9.8 ml of formamide), to each termination mix. After incubation for 2 min at 75°C, the samples were ready for gel loading. Usually 3 µ l were adequate for loading.

Gel loading and electrophoresis.

A 3 µl of the final reaction mixtures were loaded on the gel. The mixtures were run in parallel in the order T,C,G,A, on a 14cm x 80cm long, 6% (19:1 acrylamide/bisacrylamide, 1x TBE, 8M urea) sequencing gel, at constant 40 Watts. Three separate runs were made for each sequencing reaction, to enable reading of as much sequence as possible.

First run: Until the bromophenol blue had reached the bottom,

Second run: Until the xylene-cyanole had reached the bottom

Third run: Overnight, for approximately 16 hours.

CHAPTER 3
3. ATTEMPTS TO CLONE CYTOCHROME C GENES FROM *BACILLUS LICHENIFORMIS*

3.1. OBJECTIVES

The objectives of this part of the project were to clone, and thereby obtain the gene sequence(s) of cytochrome(s) c from *B. licheniformis,* for structural and evolutionary analysis

The starting point for the project was the work of Woolley (1984), in this laboratory. He demonstrated, that levels of cytochromes c were raised under conditions of low aeration growth in the presence of nitrate, although they were still comparatively low. Subsequently, he managed to purify and characterise three different c-type cytochromes; c551, c552 and c554 from *B. licheniformis,* and partially sequence the c552 cytochrome. The partial sequence of cytochrome c552 showed certain likeness to a class I cytochrome of subclass ID ("*Pseudomonas* c-551" type cytochrome). At that time, there was no sequence information on cytochromes in Gram positive bacteria and the apparent presence of this cytochrome c type in the Gram positive *B. licheniformis* came as a surprise because of the evolutionary distance between *Bacillus licheniformis* and the species previously shown to have it.

Woolley only observed single bands for the trypsin released cytochromes, by SDS-PAGE gel electrophoresis. Therefore he drew the conclusion that they were not integral membrane proteins (Woolley, 1984).

During the course of the present work, the amino acid sequencing of the partial sequence of the cytochrome c552 was completed by van Beeumen (1987, personal communication). He also partially sequenced the cytochrome c551 (62 residues) that turned out to be of the same type as the c552 cytochrome c with 32% identity. A complete gene sequence of a cytochrome c550 from *Bacillus subtilis* was also published (Von Wachenfeldt and Hederstedt, 1990). It had 70% identity with the c551

cytochrome from *Bacillus licheniformis* and was shown to be membrane bound. This was supported by analysis of the sequence. The gene can be divided into a haem containing functional part and N-terminal hydrophobic membrane attachment part. This contrasts with Woolleys' conclusion on the cytochromes in *B. licheniformis.*

3.2. CLONING STRATEGY

The strategy of cloning can be divided broadly into the steps outlined below. More detailed description is given in the appropriate sections.

1. Designing probes, on the basis of partial amino acid sequence of the c552 cytochrome of *B. licheniformis.*

2. Isolation of genomic DNA from *B. licheniformis* and estimation of the efficiency of different restriction enzymes in restricting the DNA.

3. Estimation of probe specificity for its target sequence on genomic blots.

4. Estimation of the size of the restriction fragment containing the gene.

5. Making and screening a *Bacillus licheniformis* genomic library.

a) A primary screening of the library with a high specificity probe and collecting a number of positive clones.

b) A secondary screening. The right clone would be identified by a consensus of results obtained by probing with different probes and the size of the recombinant fragment.

6. Transfer of the putative gene containing fragment to a sequencing vector and its subsequent sequencing

3.3. PROBES

Four mixed probes and one single long probe were designed and made on the basis of the amino acid sequence information on the cytochrome c552 (Woolley, 1984). They were designated 20.1, 17.1. 17.2 and 15.1, all mixed probes, and 51.1, a single long probe. Figure 3.1 shows the amino acid sequence of the cytochrome c552 from *B. licheniformis*, and the location of the probes. Figure 3.2 shows the partial amino acid sequence of the cytochrome c551 from *B. licheniformis* and the location of the 51.1 probe.

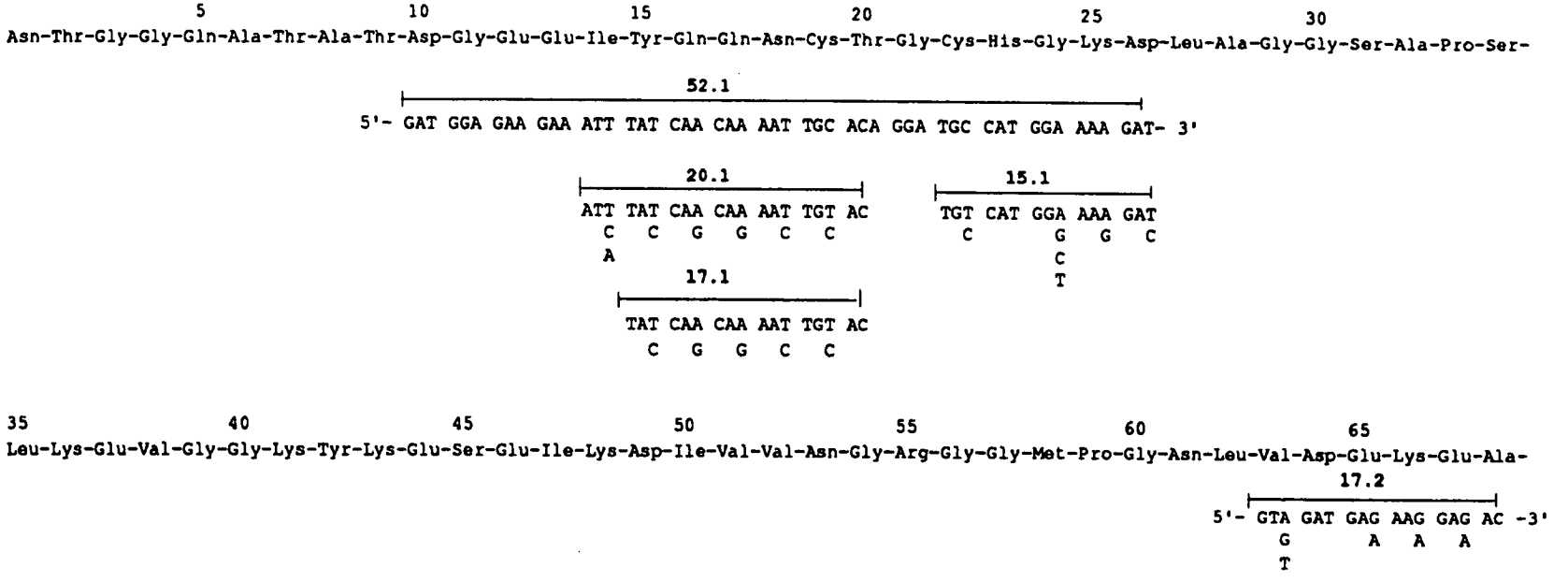


Figure 3.1 The partial amino acid sequence of c552 from *Bacillus licheniformis* and location of oligonucleotide probes

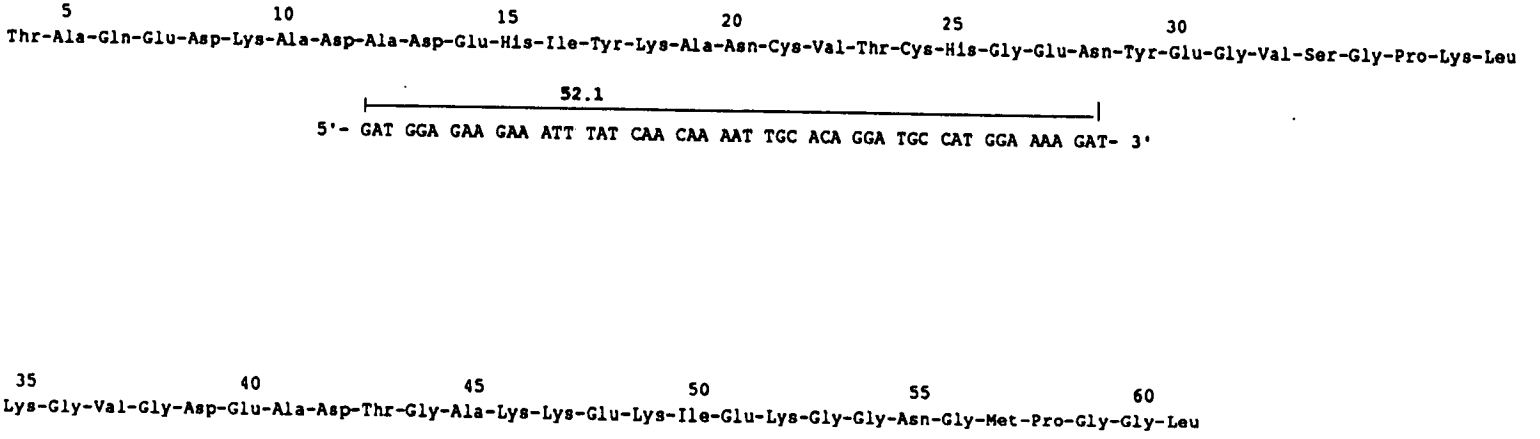


Figure 3.2 The partial amino acid sequence of o551 from *Bacillus licheniformis* and location of the 51.1 probe.

3.3.1. Probe 20.1

Sequence length=20 bases. Complexity=96. Td min.= 48°C.
GC percentage=20-50%. Mean G+C%=35%.

```
5'- ATT TAT CAA CAA AAT TGT AC-3'
      C   C   G   G   C   C
      A
```

3.3.2. Probe 17.1

Sequence length=17 bases. Complexity=32. Td min:=42°C.
GC percentage=23-53%. Mean G+C%=38%.

```
5'- TAT CAA CAA AAT TGT AC -3'
      C   G   G   C   C
```

3.3.3. Probe 17.2

Sequence length=17 bases. Complexity=24. Td min.=44°C.
GC percentage=30-53%. Mean G+C%=41.5%.

```
5'- GTA GAT GAG AAG GAG AC -3'
      G       A   A   A
      T
```

This probe was biased for the first two codons. The choice of codons was made on the basis of codon usage as calculated from three sequenced genes from *B. licheniformis*. These genes code for three extra cellular proteins: α-amylase, β -lactamase, and spoOH-protein. It was also assumed that mismatches at the end of the oligonucleotide (if occurring) would have less influence on the hybrid than mismatches in the middle.

Codon usage for amino acids coded for by probe **17.2** :

Val	Asp	Glu	Lys
GTG 0.43	GAT 0.61	GAA 0.63	AAA 0.64
GTA 0.21	GAC 0.39	GAG 0.27	AAG 0.36
GTT 0.21			
GTC 0.14			

After this probe was made, another gene coding for the extra cellular enzyme subtilisin, was sequenced. This altered the codon usage table.

Val	Asp	Glu	Lys
GTG 0.30	GAT 0.55	GAA 0.63	AAA 0.66
GTA 0.23	GAC 0.45	GAG 0.27	AAG 0.34
GTT 0.23			
GTC 0.23			

Using this table, the probability of getting the following mismatches are: Two mismatches: 0.23 x 0.45 = 0.10
One mismatch: (0.23 x 0.55) + (0.76 x 0.45) = 0.47
No mismatch: 0.76 x 0.55 = 0.42
Only one or no mismatch: 0.47+ 0.42 = 0.89

3.3.4. Probe 15.1

Sequence length=15 bases. Complexity=32. Td min for hybrids: 40°C. G+C percentage=33-60%. Mean G+C%=46%.

```
5'- TGT CAT GGA AAA GAT- 3'
      C       G   G   C
              C
              T
```

This probe was biased for the histidine codon. The probability that this codon is CAT is 0.82 according to codon usage.

3.3.5. Probe 51.1

The choice of codons was based on the codon usage in genes of four extra cellular proteins from *B. licheniformis*; α -amylase, β-lactamase, spoH-protein and subtilisin.

Sequence length=51 bases. Complexity=1.
G+C percentage=35%

5'- GATGGAGAAGAAATTTATCAACAAAATTGCACAGGATGCCATGGAAAAGAT- 3'

3.4 DNA ISOLATION

DNA was isolated from *B. licheniformis*, strain 6346/c. Electrophoresis on 1% agarose gel, confirmed that it was undegraded and virtually RNA free. The DNA was diluted to give a concentration of 1 mg/ml.

3.5. EFFICIENCY OF RESTRICTION ENZYMES

Restriction enzymes were compared for their efficiency in restricting genomic DNA from *B. licheniformis*. The following restriction enzymes were used, all of which recognise six base sequences and give staggered ends: *Hind* III, *Eco* RI, *Aat* II, *Pst I*, *Sal* I, *Pvu* I, *Xho* I and *Bam*HI. In these initial experiment, three general (low salt, medium salt and high salt) buffers were used, depending on the individual enzymes requirements.

Only *Hind*III restricted the DNA satisfactorily. Consequently, experiments were limited to the use of this enzyme and it was the only one used throughout the work with the mixed probes.

As another efficient restriction enzyme was later needed when employing the single long probe, the above experiments were repeated with modifications to see if technical improvements could be made, resulting in better and more efficient restrictions. During these attempts, care was taken to optimise conditions,

1) by using the prescribed buffers for every single enzyme and using spermidine (5mM) and BSA (0.01%) in the reaction mixtures

2) by diluting the DNA in the reaction mixtures (to 0.25 μ g/μl), to minimise the risk of inhibiting contaminants, which might have been co-purified with the DNA, and

3) by increasing restriction time and adding another aliquot of enzyme to the reaction mixture after predetermined time, which depended on how labile the enzymes were known to be. The above enzymes were tried again, but this time they were from another source in case ones used before were destabilised. In addition, three new enzymes were tried; *Bgl*II, *Xba*I and *Cla*I, which also recognise six base sequences. The results were much the same (figure 3.7). However, one of the three new enzymes, *Cla* I, cut the DNA satisfactorily and was comparable with the *Hind*III in efficiency. As

only two of the enzymes cut the DNA completely, *Hind*III and *Cla*I, the subsequent experiments were restricted to their use. Both these enzymes were efficient in the One-Phor-All buffer which became available during course of this work. This buffer was used from then on.

3.6. ATTEMPTS TO DETECT THE CYTOCHROME C GENE(S) FROM *B. LICHENIFORMIS*

Two different screening strategies were employed for detecting the cytochrome c gene(s). Both were based on the use of oligonucleotide probes, designed on the basis of the amino acid sequence of cytochrome c552 from *B. licheniformis*. One involved the use of short mixed oligonucleotide probes, but the other was based on a single long probe in conjunction with washing of hybridization filters with tetramethyl ammonium chloride solution.

3.6.1. The Mixed Probe Strategy

Two or more, short, mixed oligonucleotide probes are made from two different regions in the protein, preferably far apart. Genomic blots restricted with different restriction enzymes are subsequently probed and the specificity of the probes evaluated by washing the hybridisation filters at increasing stringency (increasing temperature). If the specificity of the probes turns out to be slight and no reliable discrimination is obtained between two or more bands it is expected that a consensus of two probes will ensure the identification of the correct band, provided the restriction enzyme used for the fragmentation of the DNA will not cut between the two target sequences. To guard against this latter possibility, genomic blots from two or more different restriction enzymes should be probed in parallel. The probes were labelled with radioactive phosphorous (γ ^{32}P-ATP) and the level of incorporation of radioactive label was estimated by the CTAB method (see chapter 2). The probe was used in the hybridisation experiments if the specific activity was above 10^5 cpm. Continuation at each stage was based on sequence information, available at each time, and outcome of preceeding experiments.

3.6.1.1. Probing genomic blots with the mixed probe 20.1

The first probe designed and used, was a 20mer in a region of the low codon redundancy, covering the sequence residues 20-27 (figure 3.1 and section 3.3.1). This is the haem site region with the corresponding invariant cysteine and histidine residues. The probe pool consisted of 96 sequences, including all possible codon choices for each amino acid. Other regions were not considered suitable because of a high redundancy, but at the time, sequence information was limited.

Genomic DNA (10-15 μg) was restricted with *Hind*III and electrophoresed in 1% agarose gel. The DNA was transferred onto cellulose filters by Southern blotting and probed with the **20.1**, probe which had been labelled as described in methods. Hybridisations were performed at 37°C. The starting temperature for the washing of the filters was 20°C. No hybridisation was observed despite repeated trials.

3.6.1.2. Probing with the mixed probe 17.1

The inability of the **20.1** to hybridise and give distinct bands was unexpected. Apart from some undetected mistake in the execution of the experiment, a possible explanation was thought to be the combined effect of the relative low G+C percentage (mean G+C%=35%) of the probe, and of its high complexity (98). The latter might have made the concentration of probe-sequences too low, giving exact and near to exact match. Therefore, a shorter probe, **17.1**, a 17mer of less complexity (32), was designed from the same region as the **20.1** probe, including the first cysteine codon of the haem site.

Experiments were conducted as described above, on a *Hind*III restricted genomic DNA. Hybridisation was done at 30°C as well and washing was started at 20°C. As before, no hybridisation was observed.

3.6.1.3. Probing with the mixed probe 17.2

At this time, more sequence information on the c552 gene had become available. It was therefore decided to make a third probe from a different region and aim at still lower complexity and higher G+C percentage. This new probe, **17.2**, was designed from a region near the C-terminal part of the protein and was a 17mer, biased relative to two codons. It had a complexity of 24 and mean G+C percentage of 41.5%.

Experiments were conducted as before, with hybridisation done at 37°C and washing started at 20°C. At the lowest stringency washing, nonspecific binding to a high degree was observed. At medium stringency

no less than four distinct bands were visible on the auto radiographs. According to the marker (λ CI 857 *Hind*III restricted), these bands are approximately 6.5 kB, 4.4 kB, 1.9 kB and 1.7 kB. No real discrimination could be made between the four bands, but possibly the biggest and the smallest ones were a grade more intense than the two smallest ones. Figure 3.3 shows the result of a similar experiment.

3.6.1.4. Probing with the mixed probe 15.1

The results of probing with the **17.2** probe were not conclusive, as a single band was not observed. A different probe was therefore considered to be necessary to distinguish between the bands. The results of the probings described above, indicated, that there could be a relationship between the ability of probes to bind to the genomic blots and the complexity and/or high G+C percentage of the probes. Therefore, a 15mer, designated **15.1**, was designed of low complexity (32) and with a high G+C percentage of 46%. This probe included the codons for the second cysteine and the histidine of the haem site.

DNA from two *B. licheniformis* strains was used in subsequent experiments, the strain 6346/C from which the cytochrome c had been isolated from, and a strain designated 749. DNA from each strain was restricted with *Hind*III and 10 μ g samples of DNA from each strain were loaded, side by side on 1% agarose gel in duplicate and electrophoresed over night. The DNA was transferred onto nitrocellulose filters and the filter cut into identical strips. The strips were each probed with a different probe; the **15.1**, and **17.1**. In another experiment the genomic DNA was probed with the **15.1** and the **17.2**.

It was not until the hybridisations were performed at 20°C, well below the minimum Td of all probes that all the probes bound to the DNA. Figure 3.3 shows the results of probing blots of genomic DNA from *B. licheniformis* with the probes **17.2** and **15.1**. Figure 3.4. shows the results of probing with the probes **15.1** and **17.1**.

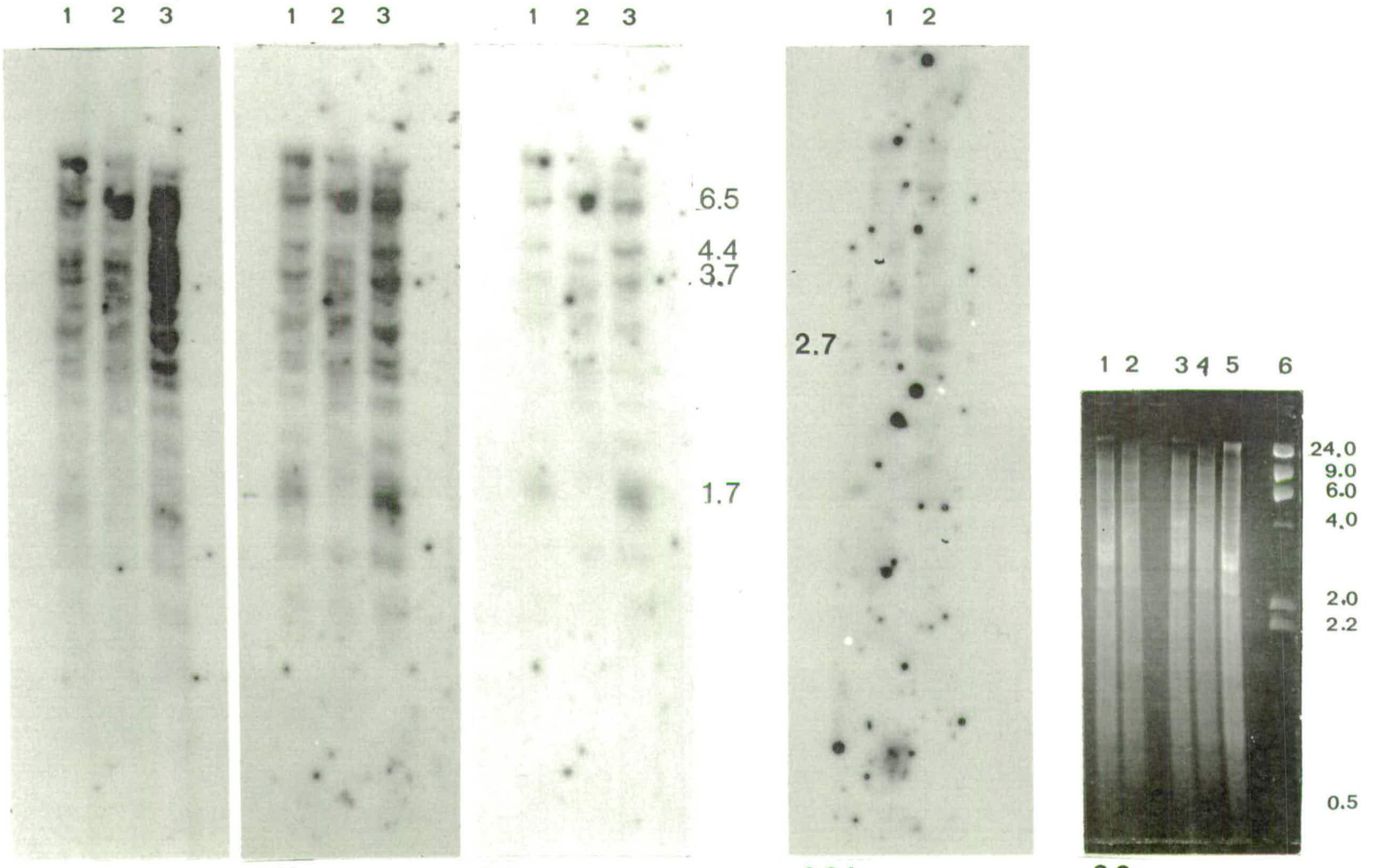
There appears to be a distinct difference in the ability of the probes to bind to the genomic blots. They can be ordered in descending order on the basis of the signal strength of bands **17.2 > 15.1 > 17.1**. This difference between them could also been seen in the degree of unspecified binding to the DNA, and was also reflected in the number of background spots on the filters. At the most stringent wash, 35°C (5 min and 120 hours exposure) the **15.1** and **17.1** both gave a single strong band of seemingly the same size (2.7 kB), indicating that they had bound to the same fragment, and

Figure 3.3. Hybridisation of the mixed probes, 17.2 and 15.1 to blots of *Hind*III restricted genomic DNA from *B. licheniformis*.

Figure 3.2a-3.3c. Results of probing with the 17.2 probe after washing at different temperature stringencies, 20°C, 37°C, and 43°C. Lane 1 & Lane 3: *B. licheniformis* NCIB 9346; Lane 2: *B. licheniformis* 749.

Figure 3.3.d. The results of probing with the 15.1 probe after washing at 34°C. Lane 1: *B. licheniformis* NCIB 6346. Lane 2: *B. licheniformis* 749.

Figure 3.3e. The corresponding ethidium bromide stained agarose gel from which the DNA was transferred. Lanes 1 and 2 correspond to lanes 1 and 2 in figure 3.2d. Lanes 3-5 correspond to lanes 1-3 in figure 3.2a-c. Lane 6: Marker DNA *Hind*III restricted λIC I857.



3.3a. 3.3b. 3.3c. 3.3d. 3.3e.

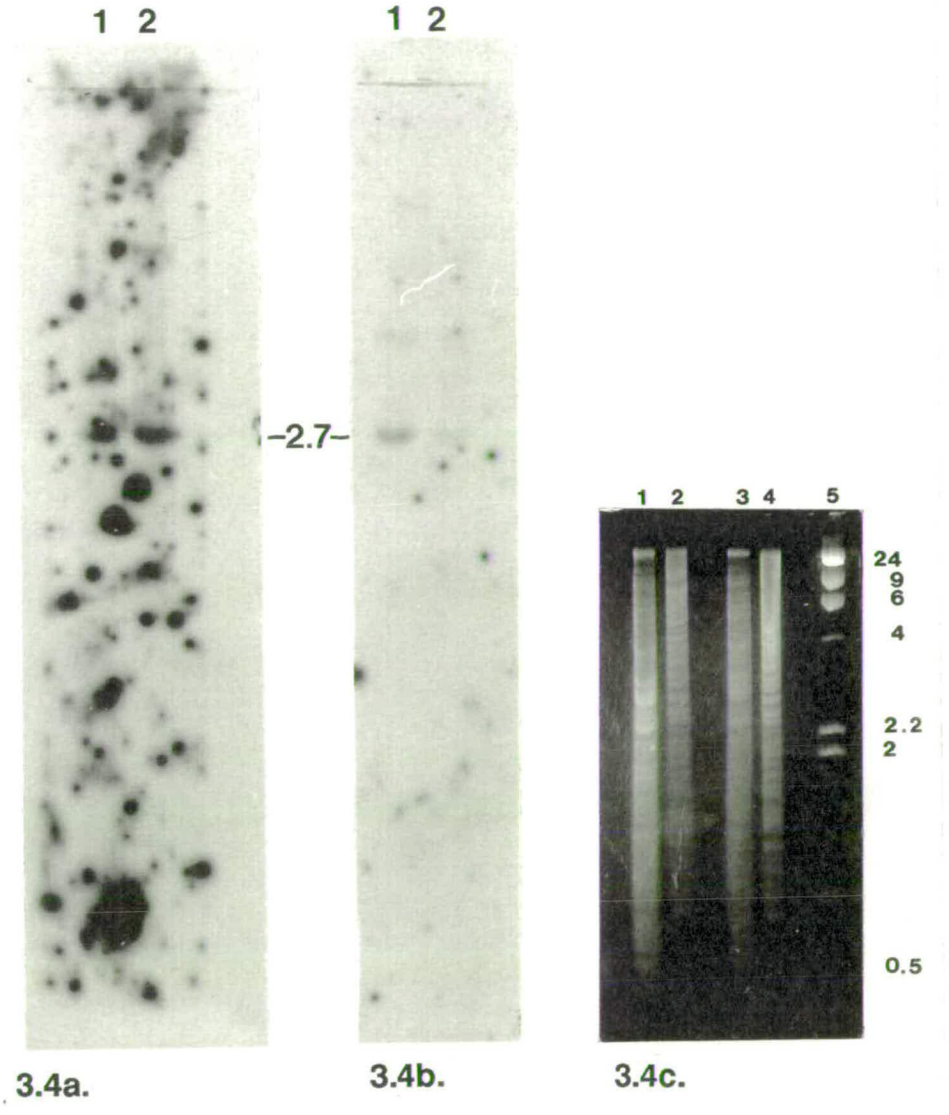


Figure 3.4. Hybridisation of the mixed probes 17.1 and 15.1 to blots of *Hind* III restricted genomic DNA from *B. licheniformis*
Figure 3.4a. The results of probing with the 17.1 probe after washing at 35°C.
Lane 1: *B. licheniformis* NCIB 6436;
Lane 2: *B. licheniformis* 749.
Figure 3.4b. The results of probing with the 15.1 probe after washing at 35°C.
Lane 1: *B. licheniformis* NCIB 6436.
Lane 2: *B. licheniformis* 749.
Figure 3.4c. The corresponding ethidium bromide stained agarose gel from which the DNA was transferred. Lanes 1 and 2 correspond to lanes 1 and 2 in figure 3.4a&d. Lanes 3 and 4 correspond to lanes 1 and 2 in figure 3.4b. Lane 5: Marker DNA *Hind* III restricted lIC 1857.

therefore to the gene itself. It is notable that this temperature is not so far below the melting temperature of **15.1** which is 40°C. The **17.2** probe does not appear to bind to this fragment.

3.6.1.5. Comments

There are apparently inherent differences between the ability of the probes to bind to the DNA. The optimum temperature, (empirically determined for mixed oligonucleotide probes to selectively hybridise to their target sequences), is calculated to be 5°C lower than the melting temperature (Td) of the probe. In the initial probing experiments, hybridisation were done at 37°C, except for probe **17.1** where the temperture 30°C was used as well. This is more than 5°C lower temperature than the minimum melting temperatures (Td). Only probe **17.2** bound to the genomic blots. Later, when hybridisation were done at much lower temperature, (20°C) all three probes, **15.1**, **17.1** and **17.2** bound to the DNA blots.

The probe **17.2** bound more strongly than the other two and gave a greater number of, and more intense bands, but also a higher background hybridisation. The **17.1** probe gave by far, the weakest binding. Notable was, that the **15.1** probe gave a markedly higher background hybridisation than the **17.1** probe.

The procedure in the hybridistion experiments was the same in all aspects, such as concentration of probe and other reagents. The level of radioactive label incorporated by the probes, was usually estimated by the CTAB test (see chapter 2) and found to be similar and of sufficiently high degree.

By examining the probes, it is unlikely that the difference in their binding capability is because of complementary secondary annealing. Rather, it seems be the result of combination of physical attributes; the G+C percentage, length of the probe, and possibly the complexity of the probe. The probes can be ordered in a descending order, according to their ability to hybridise to the DNA; **17.2** > **15.1** > **17.1** (>**20.1** not used in the last trials). The **20.1** probe has the lowest mean G+C percentage, and highest complexity. It did not hybridise at all (hybridisation was only done at 37°C). The **17.2** probe has markedly higher mean G+C percentage and gives the strongest signal, even stronger than the **15.1** probe which has the highest G+C percentage. The difference between the **17.2** and the **15.1** probes may be explained by the higher complexity, smaller size of the **15.1**, and consequently, a lower melting temperature.

The probes 15.1 and 17.1 gave a common single strong band in the *Hind*III restricted genomic blot. Apparently the 17.2 probe does not bind to this fragment. This indicates either that the probe does not match its target sequence or that there is a *Hind*III restriction site between the target sequences of 17.2 probe and the 15.1 and 17.1 probes. In fact, there is a possible *Hind*III restriction site at position 33 in the amino acid sequence, involving the codons of 33 pro ser leu 35. The possibility of this *Hind*III restriction site, out of all possible codon combinations for these three particular amino acids, is very low, little less than 1 out of 72, assuming total randomness in codon usage, and a little less than 1 out of 53, using most probable codon usage.

It was concluded, that screening libraries with 15.1 might be feasible, but it was decided to follow a different course by using a long single probe. The single probe approach was expected to be more effective and more versatile (see below).

3.6.2. The Single Long Probe Strategy

This strategy was devised, as more conclusive results were desired than had been obtained with the mixed probe approach. The specific aims were:

1) To increase the discrimination in the screening for the c552 cytochrome gene on which the probe was based and

2) To screen for related cytochromes.

3.6.2.1. Theory

Any single probe, based on amino acid sequence, has a precise minimum homology of 66.6% (assuming that the third position in the probe is wrong in all codons), and to that can be added the expected percentage of all nucleotides in the third position, to be right. The predicted homology should be 74%, assuming random choice of base into the third position, or greater, if codon preference is taken into account. (Lathe, 1985).

The effect of complexity is minimal as there is only one sequence, and therefore, an increased probe length confers selectivity. In contrast, the complexity of a mixed probe increases fast by increasing length and consequently their size is limited. For a large size mixed probe he concentration of probe needs to be increased (to unrealistic levels) in hybridisation experiments to compensate for the smaller fraction of the

pool giving exact (or near exact) match. Above a certain threshold there is an increase in background spots, which will mask any bands, and the increased number of different sequences in the pool increases the likelihood of fortuitous matching.

The contribution of tetra methyl ammonium chloride to increase selectivity of probing with a single long probe, is in eliminating the dependence of the melting temperature, Td, on the G+C content. Tetraalkyl ammonium salts bind to A-T base pairs and have been used to abolish the preferential melting of A-T versus G-C. With this selective binding of Me_4NCl, the dissociation equilibrium is displaced, raising the melting temperature. At 3M concentration this displacement is sufficient to shift the melting temperature of A-T pairs to that of G-C base pairs. This results in a sharpening of the melting profile so that DNA hybrids, which melt over a range of 5-10°C, melt within 1°C range in Me_4NCl (Wood *et al.*, 1987).

The advantage of using this method lies in the fact that the difference between melting temperatures of different hybrids lies in the the number of exact matches and therefore the selectivity is increased by increasing probe length. The number of false positives is expected to fall more sharply with a gradual increase in washing stringency than when short mixed probes are used. The number of hybrids observed close to the conditions of highest stringency (where only a single signal is observed), are expected to be fewer and are more likely to be related to the target sequence.

As said above the **51.1** probe was designed on the basis of the c552 sequence and according to the available information on codon preference in extracellular enzymes from *B. licheniformis*. It has 66.6%minimum sequence homology with the c552 gene. It is designed for a region of low redundancy in the sequence, where most of the codons have only a choice of two different bases in the third position. Moreover, according to the codon usage table for *B. licheniformis* the selected third base, has average frequency of approximately 0.6.in the genes coding for the extracellular enzymes. The predicted overall homology probe-cytochrome c552 therefore increases to 84% .

The **51.1** probe has a minimum homology of 43% (22 identities) and a maximum homology of 70% (max. 37 identities) with the *B. licheniformis* cytochrome c551 gene. It has 14 positions which will always be wrong. There are 15 base positions in the sequence, with an 0.55 average frequency of occurence in the genes, besides the 22 base positions which

will always match the probe sequence. The predicted overall homology probe 51.1-cytochrome c551 gene is therefore 60%. It is a fair chance that this probe will pick the c551 gene at lower stringency than is optimal for the c552 gene. A further help is in the fact, that the the "invariant" and possible mismatches are not evenly distributed over the sequence. Homology between the two cytochromes is greatest in the region which corresponds to the middle right part of the probe sequence, but falls off towards the ends.

3.6.2.2. Results

51 bases long oligonucleotide probe was designed, on the basis of partial sequence of the **c552** cytochrome, from a low redundancy region including the haem site (figure 3.1, section 3.3.5). The target sequence of this probe included the target sequences of the probes **20.1**, **17.1**, and **15.1**. The expected overall match for this probe to its target sequence was 84%. This probe is expected to have a Td of 60°C when 6xSSC washing buffer is used. This Td is equivalent to that of a perfect 30mer (the Td increases considerably when Me_4NCl is used.) The frequency of a chance occurrence of a sequence, 26 bases long and homologous to the probe, is 0.001 for a bank of 10^7 nucleotides (the bacterial genome is around 10^6 nucleotides) (Lathe, 1985).

The genomic DNA, was restricted with *Hind*III, *Cla*I and *Eco*R. A single strong band was observed in all restriction blots from washing temperature of about 55°C. The sizes of DNA fragments were 2.7 in *Hind*III restricted genomic blots, 2.9 in *Cla*I restrictions, and bigger than 20 Kb in the *Eco*RI restrictions (figure 3.5).

3.7. ATTEMPTS TO CLONE CYTOCHROME C GENE(S) FROM *B. LICHENIFORMIS*

3.7.1. The Mixed Probe Approach

3.7.1.1. The screening of libraries with short mixed probes

These experiments were done at the stage when only the 17.2 probe hybridised to the genomic DNA blots. The aim was to obtain clones which would be available for analysis by other probes still to be obtained. Genomic DNA was restricted with *Hind*III and cloned into the selection vector λ 1149, also restricted with *Hind*III. *E. coli* cells were transformed

with unpacked recombinant lambda DNA. 2500 clones were screened. Eight positive clones were picked and analysed.

The DNA from these eight positive clones was restricted with *Hind*III. One clone had an insert that could not be cut from the vector. Two had an insert of the same size, 1.7 kB. The other clones had inserts of the sizes 3.7 kB, 2.9 kB, 2.8 kB, 1.9 kB, and 1.7 kB. One of these clones, L67, had furthermore, beside the 1.7 kB band, two other inserts that had been cointegrated into the vector (of approximately 0.5 kB and 2.0 kB).

By comparing these inserts with *Hind*III restricted genomic blot of *B. licheniformis*, probed with the 17.2 probe, three of the inserts correspond to bands on the blot; the 1.7 kB insert of L2 and L67, which was one of the strong bands, the 3.3 kB insert and the 2.7 kB insert of two other clones. The two latter inserts correspond to rather faint bands of the genomic blots, one of which is not visible after the more stringent washes.

By probing the DNA from the clones with probe 17.2, and doing increasingly stringent washes, it was found that the 1.7 kB fragment from either L2 or L67, hybridised most strongly to the 17.2. As neither probe 15.1 nor 17.1 hybridised to any of the clones, further work was deemed rather futile.

3.7.1.2. Sequencing of the 1.7 kB L67 fragment

As noted above, there is reason to believe that the *Hind*III cuts inside the gene. Therefore, the C-terminal part of the gene may be at either end of the 1.7 kB fragment. An attempt was made to sequence the fragment from both ends.

The L67 fragment was cloned directly into the M13mp18 sequencing vector. The L67 clone, which has three *Hind*III fragments, was restricted with *Hind*III and also the M13 sequencing vector. The DNA from the phage was then ligated directly into the sequencing vector *E.coli* NM522*. Clones were obtained by transformation of this *E. coli* strain which is repressive of Lambda Phage expression. A total of 45 of these clones were collected, 19 of which contained the right fragment. They were probed with the 17.2 probe to examine the orientation of the fragment. All had the same orientation and did not bind to the probe. The fragment in this orientation was sequenced with the standard Sanger method. 280 bases long sequence was obtained. It bore no relation to the cytochrome c552 sequence.

3.7.2. The Single Long Probe Approach

3.7.2.1 The screening of library with the 51.1 probe

Two strategies were devised, as outlined below:

1) Construction of a library from a size selected *Cla*I restricted DNA, using a Mp 13 derived cloning vector MP L30, and screening for the c552 cytochrome gene only. The fragment range was to be chosen on the basis of probing *Cla*I genomic blots. DNA was selected between 0.5 kB above and below the fragment size (2.7 kB) to which the probe bound to.

2) The screening of a complete genomic library in the λ (1149) vector, for the c552 gene cytochrome c gene, and the collection of a number of clones (putative c551 clones) at lower stringency.

3.7.2.2. The use of *Cla* I restricted, size selected DNA in attempts to clone the cytochrome c 552 into the MP L30 vector

As there were reasons to believe that a *Hind*III restriction site was in the cytochrome c552 gene, it was decided to use *Cla*I restricted DNA in these cloning attempts. By using size selected DNA, the number of clones needed to be screened, is reduced at least tenfold. The fragment believed to contain the gene is 2.9 kB (see figure 3.4). This is a suitable size for cloning and the subsequent sequencing. The *Cla*I restricted DNA, 150 μg was electrophoresed in a 1% agarose gel overnight, as described in methods.

Two different methods were used to elute the DNA of desired size range (between 3.4 kB and 2.4 kB) from the gel. The Geneclean extraction method was used and fragments of the desired fragment length were electrophoresed into DEAE 52 paper. Figures 3.6a and 3.6b. show an electrophoretic analysis of the size selected DNA obtained by these methods.

Incorporation of DNA into the *Cla*I site in the MP L30 vector is supposed to disrupt the expression of the beta galacosidase gene. Therefore, recombinant plaques can be recognised by their white colour (wild type plaques are blue, see section 2.1.3.).

Despite repeated attempts, ligation of the size selective DNA into the λ 1149 cloning vector was unsuccessful. Controls showed that ligation was working. The restricted plasmid religated and transformation was also working. No incorporation of genomic DNA was observed, despite that the size selected DNA had been prepared with two different methods to rule out any effect by the purification method on the ligation.

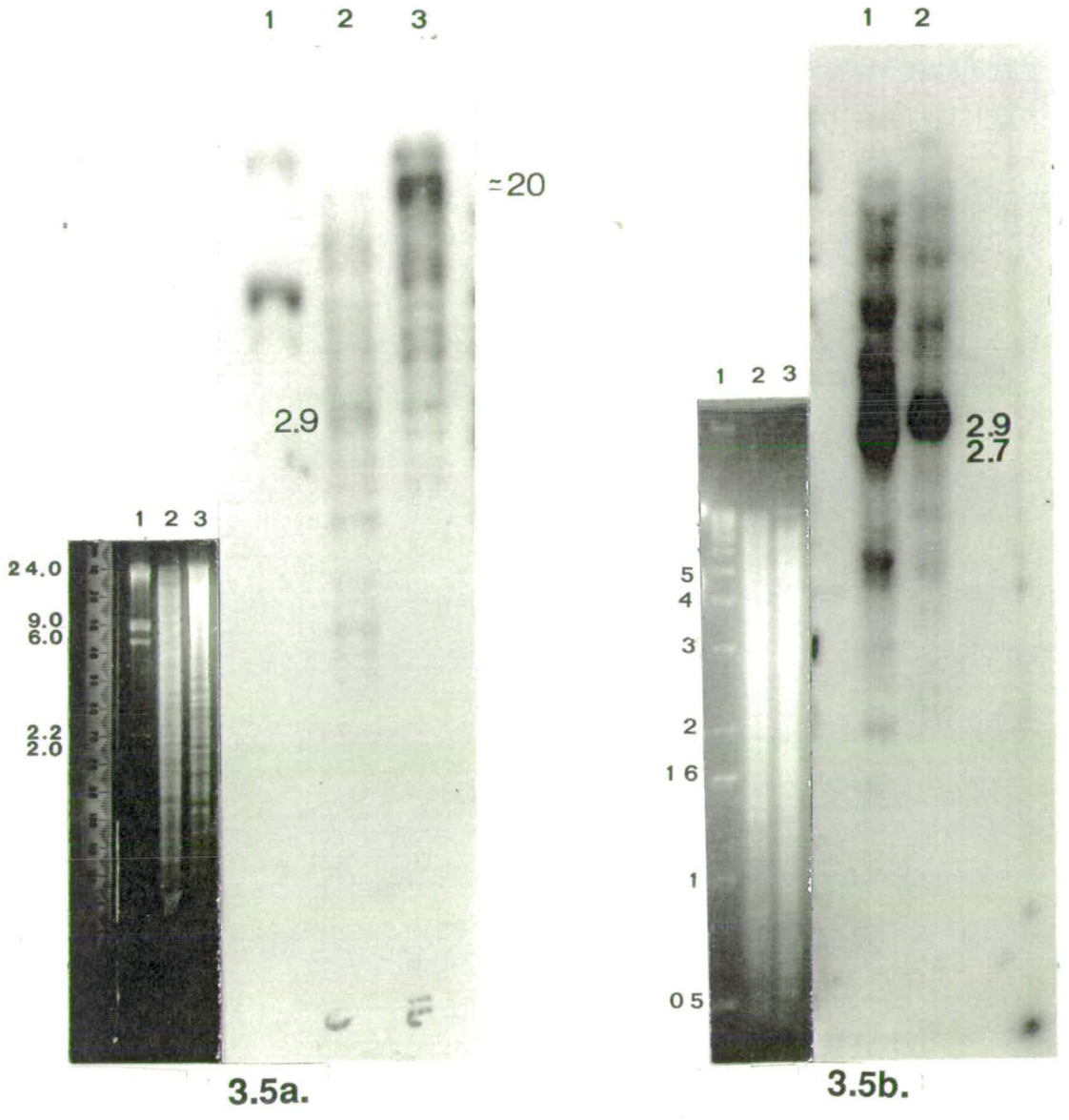


Figure 3.5. Hybridisation of the single long 51mer probe to blots of restricted genomic DNA from *B. licheniformis* NCIB 6436.
Shown are two radiograms from two different experiments and the corresponding ethidium bromide stained agarose gels from which the DNA was transferred.
Figure 3.5a. Lane 1: Markers. Lane 2: *Cla*I restriction. Lane 3: *Eco* RI restriction.
Figure 3.5b. Ethidium bromide stained agarose gel: Lane 1: Markers. Lane 2: *Hind* III restriction. Lane 3: *Cla*I restriction. *Radiogram* Lane 1: *Hind* III restriction. Lane 2: *Cla*I restriction.

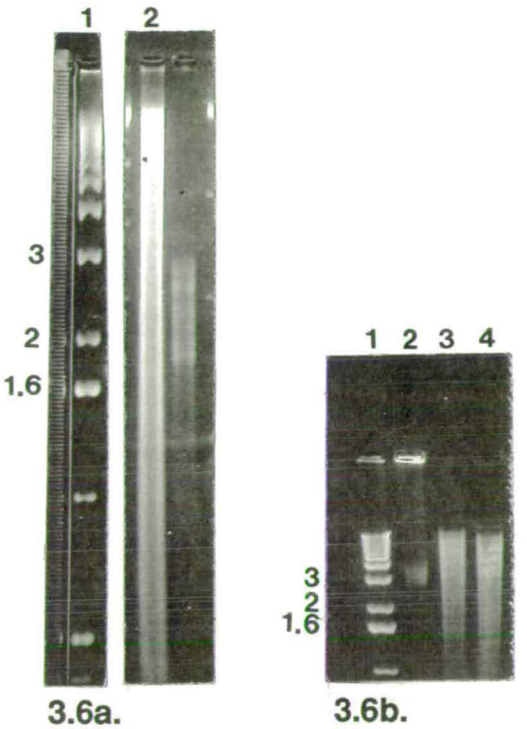


Figure 3.6. Electrophoretic analyses of size selected *B. licheniformis* genomic DNA
Figure 3.6a. Electrophoresis of size selected DNA after extraction from the agarose gel by the Geneclean method.
Figure 3.6b. Electrophoresis of size selected DNA after electrophoresis out of the agarose gel by the DEAE filter method.

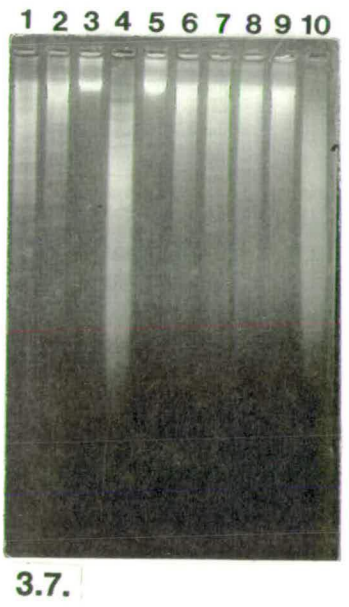


Figure 3.7. Examination of the efficiency of different restriction enzymes in restricting genomic DNA from B. licheniformis.
Lane 1: *Sal*I; *Lane 2*: *Eco*RI; *Lane 3*: *Xba*I; *Lane 4*: *Hind*III;
Lane 5: *Pst*I; *Lane 6*: *Pvu*I; *Lane 7*: *Bgl*II; *Lane 8*: *Xho*I;
Lane 9: *Bam*HI; *Lane 10*: *Cla*I.

Time did not allow further experiments by this approach. In retrospect, the only explanation I can venture is that the blue and white response was not effective so that recombinant plasmids were not recognised. This is very unlikely, but could have been checked by restriction analysis of a number of phages from the transformation. Unfortunately, this was not done.

3.7.2.3 The use of long single probe for screening λ 1149 genomic library for cytochrome c genes

3.7.2.3.1. Construction of genomic library in λ1149

A complete genomic library was constructed in λ1149. DNA was restricted with *Hind*III and ligated into the vector which had also been restricted with *Hind*III. The in vitro packaging extracts of Amersham were used for packaging the recombinant phage DNA.

3.7.2.3.2. Screening

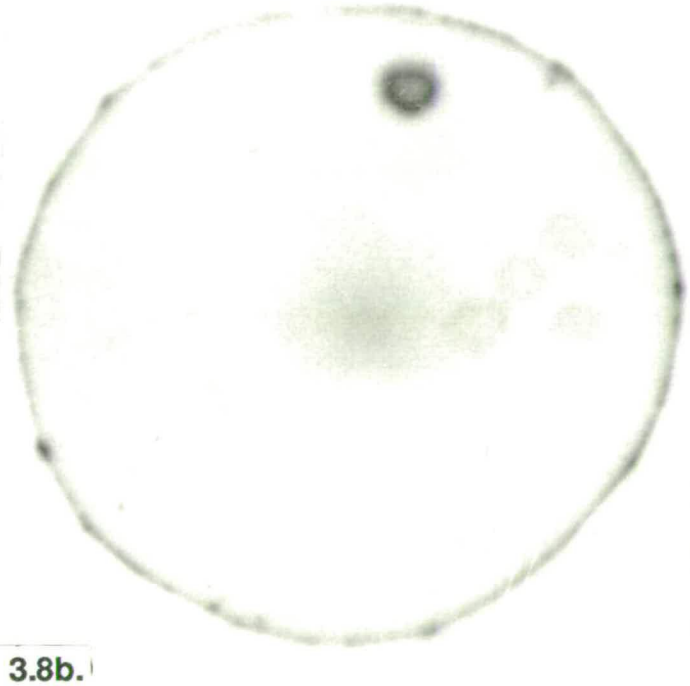
An *E. coli* strain, which only allows growth of recombinant phages, was transformed and the cells plated out on to a 22x22 agar plate, to give an average density of 10 plaques/cm^2. This gave close to 4000 plaques. The plaques were replica blotted onto a H-bond nylon filter, in duplicate, and both filters were probed with the radiolabelled 51mer. Hybridisations were done at 37°C and washings were started at 20°C.

At 50°C stringency, 30 positives were collected. They were purified and plated for secondary screening onto a grid on an agar plate. These were re-screened with the 51mer. One of the plaques gave by far the strongest signal and was only washed off at approximately 70°C (figure 3.8). This clone was designated L51.

3.8.2.3.3. Analysis of the L51 clone

DNA from clone L51 was isolated and restricted with *Hind* III. A sample of the DNA was electrophoresed in a 1% agarose gel. It was found that the phage had incorporated 4 *Hind*III fragments (see figure 3.9b, lane 3; figure 3.10, lane 4). The DNA was transferred onto a nylon filter by Southern blotting. The filter was probed with both the 51.1 and the 15.1. The 51.1 probe hybridised to the smallest fragment (figure 3.9a) but the 15.1 probe to none. According to these results, it was doubtful that the vector had incorporated the cytochrome c552 gene. The fragment size is wrong and the short mixed probes, which were specifically aimed at

3.8a.



3.8b.

Figure 3.8. Secondary screening of selected clones from a B. licheniformis genomic DNA λ-phage library of approximately 4000 clones.
Figure 3.8a. Radiogram of the results of hybridisation with the 51mer single long probe and after washing of the hybridisation filter at 55°C.
Figure 3.8b. Radiogram of the results of hybridisation with the 51mer single long probe and after washing of the hybridisation filter at 70°C. The clone which still gave a strong signal after washing at this stringency, was designated clone L51.

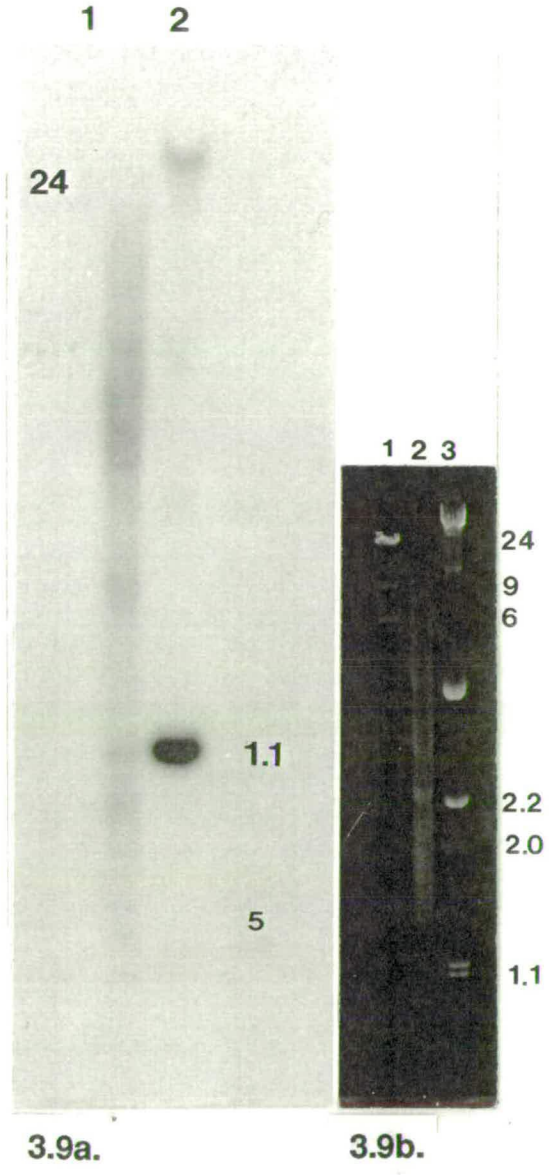
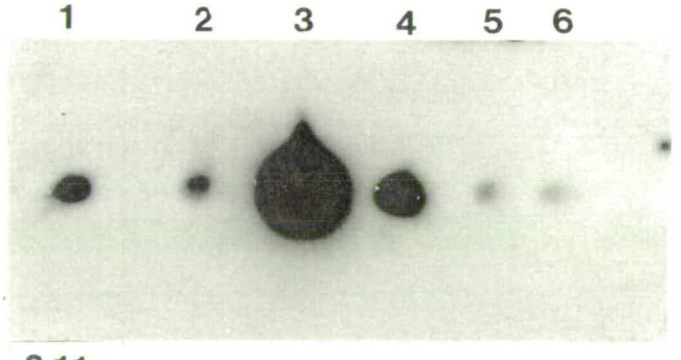


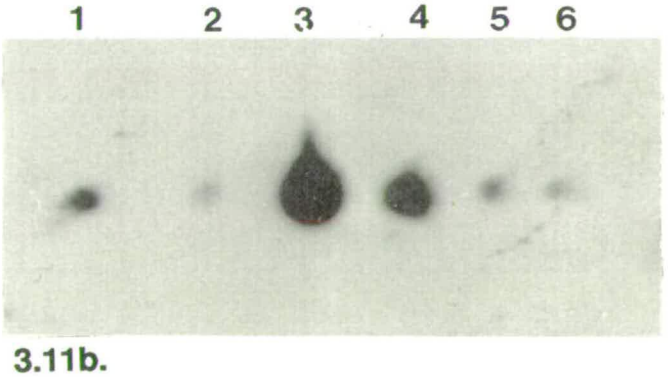
Figure 3.9. Hybridisation of the 51mer single long probe to *Hind*III restricted DNA from the putative cytochrome c clone, obtained after screening λ-phage genomic library of *B. licheniformis*.
Figure 3.9a. Radiogram of the results of hybridisation of the 51mer after washing at 70°C.
Lane 1: *Hind*III restricted genomic DNA from *B. licheniformis.*
Lane 2: *Hind*III restricted DNA from the clone L51.
Figure 3.9b. The agarose gel from which the DNA was blotted onto the hybridisation filter.
Lane 1: Marker DNA, *Hind*III restricted λIC 1857.
Lane 2: *Hind*III restricted *B. licheniformis* genomic DNA;
Lane 3: *Hind*III restricted DNA from the clone L51;

3.10.

Figure 3.10. Restriction analysis, with *Hind*III, of DNA from the L51 λ-clone and the Mp13 subclones.
Lanes 1-3: Different Mp13 subclones; Lane 4: λ-L51 clone from which the DNA fragment containing the putative cytochrome c gene was subcloned from.



3.11a.



3.11b.

Figure 3.11. Hybridisation of the long single 51mer probe to dot blots of Mp13 subclones, containing the putative cytochrome c gene.
Dots 1 and 2: DNA from different Mp13 subclones containing the putative cytochrome c gene.
Dot 3: The DNA fragment containing the putative cytochrome c gene.
Dot 4: The L51 λ-clone containing the sam fragment.
Dots 5 and 6: Dna from wild type Mp13 phages.
Figure 3.11a. The results of the probing with the 51mer after washing at 37°C.
Figure 3.11b. The results of the probing with the 51mer after washing at 60°C. The subclones (dots 1 and 2) apparently have the DNA fragment in different orientations. This was subsequently confirmed by sequencing.

hybridising to the haem site regions containing the codons for the invariant amino acids cysteines and the histidine. As the probe has a 60% predicted overall homology to the cytochrome c551 gene, the clone could have incorporated that gene, or even another related cytochrome c gene. It was therefore decided to sequence this fragment.

3.8.2.3.4. The sequencing of the L51 clone

The band, which the 51mer bound to, was electrophoresed into DEAE 52 paper, eluted and then cloned into M13mp18 vector. A number of plaques were collected. Mini preparations of DNA were made from the clones and subjected to *Hind*III restriction analysis on minigels. *Hind*III restricted L51 DNA and restricted donor phase DNA was run in parallel as a control (figure 3.10). This verified that the fragment of a correct size had been cloned into the sequencing vector. The orientation of the fragment was established by reprobing the clones with the **51.1** (figure 3.11) as both orientations were needed for fast sequencing.

800 bases were sequenced from each end by the extended, sequencing method (see chapter 2). The sequences, obtained from the opposite ends, overlapped. Complete sequence was 1300 bases. No homology to cytochrome c was found and the codons for the invariant cysteines and histidine were not found. The sequence, which came closest to match the probe had 40% identity.

3.8.2.3.5. Comments

Probe selectivity. It can be concluded from these experiments that the criteria for a sufficient selectivity are fulfilled by the 51mer probe;

1) a prominent single strong band was observed and

2) single bands were observed in blots obtained with different restriction enzymes (*Cla*I, *Eco*RI and *Hind*III). The band in the *Hind*III restriction was of the same size as obtained with the **15.1** and **17.2.**

Screening of the library. If the size of the *B. licheniformis* genome is $5x10^6$ bases, assuming complete randomness of sequence representation, and the *Hind*III cuts, on the average, every 4 kB, approximately 4000 clones are needed to give a 0.99 probability of including a particular sequence.

Close to 4000 plaques were screened, giving the approximate probability of 0.95% of including every gene coding sequence. The filters were of a good quality and gave high degree of unspecified binding at the lowest stringency temperature, and good demarcation of individual plaques. A

fair proportion of these phages must have incorporated more than one fragment. The λ 1149 is also size selective, incorporating only fragments under 11 kB, and therefore the effective size of the genome represented in the library is reduced. The number of plaques screened should have been large enough for obtaining the cytochrome genes. However, at this stage it is logical to screen a larger library before anything else is attempted.

During this work, it can be said that the questions addressed at the start of the project, could be answered, more or less, on the basis of the information made available from other sources (section 3.1), and from my work on a different *Bacillus* species (*Bacillus azotoformans*). Consequently, and also because of the difficulties encountered, this part of the project was ranked less in importance than the isolation, biochemical and structural characterisation of cytochromes c from B. *azotoformans*. It was considered that the latter could supply most of the answers, as well as expanding the scope of the study, of cytochromes c from the genus *Bacillus*.

3.9. CONTINUATION

1) PCR: As the c552 cytochrome is proposed to have an N-terminal membrane attachment part, PCR will only suffice to get the haem domain part. No information for destaining a primer is available from the N-terminal membrane attachment region. However, a PCR amplified fragment, from the haem domain could be used as a highly specific probe for the complete gene.

2) The use of a size selected DNA for making a λ-phage genomic library. The size of the library needed to be screened can be reduced by a factor of ten.

3) Direct genomic sequencing. It is possible, with a probe labelled to high specific activity, to sequence genomic DNA directly. This is done by PCR techniques. The DNA is amplified, although only linearly but the sensitivity can be increased by using size selected DNA. The direct sequencing should give approximately 200 base pairs which should be more than enough to cover the N-terminal sequence. This method should be relatively easy and quick, and is perhaps the preferred one for obtaining sequences for evolutionary and phylogenetic analysis.

CHAPTER 4
4. PURIFICATION AND CHARACTERISATION OF CYTOCHROMES C FROM *BACILLUS AZOTOFORMANS*

4.1. PRELIMINARY STUDIES

The objectives of the preliminary studies were:

1) To assess the cytochrome c content in aerobic and denitrifying cultures.

2) To compare different methods for releasing cytochromes c from cells.

3) To choose the media and conditions for large scale growth of cells.

4.1.1. The Growth of Small Scale Cultures and Cell Yields

4.1.1.1. Aerobic cultures

One litre of aerobic cultures was grown in 5 litre flasks, containing Luria-broth (L-broth). Cultures were grown overnight at 37°C, on a mechanical shaker, at a velocity which ensured maximum aeration. Inoculum was 1% from 20 ml aerobic cultures, grown from a single colony. The growth of the cultures was monitored spectrophotometrically at 650 nm. Maximum density was obtained after 12 hours. Cell lysis was not observed.

4.1.1.2. Anaerobic cultures

Small scale anaerobic cultures were grown in 500 ml media bottles, with the lid screwed on tightly, containing L-broth medium supplemented with KNO_3 (0.5%). Cultures were grown overnight at 37°C. Inoculum was 10% (to ensure immediate or early onset of anaerobic conditions), from 100 ml denitrifying cultures, grown from single colony for 24-36 hours. Maximum density was observed after 12 hours. Cell lysis was always discernible and it became more pronounced with increasing incubation time.

4.1.1.3. Harvesting of cells

Cells were usually harvested after incubation overnight, by centrifugation at 8500 g for 30 min in GS3 rotors. Denitrifying cultures were degassed by shaking, before centrifugation. Cell yields were on the average 6.83 g/l in aerobic cultures, and 5.71 g/l in anaerobic denitrifying cultures. The latter figure may be an underestimate because of cell lysis, which always occured in the denitrifying cultures.

Pelleted cells had a distinct red tinge, which was more pronounced when they had been grown under denitrifying conditions.

The red colour was interpreted as indicative of an abundance of cytochromes.

4.1.2. Cytochrome C Content and Release from Cells, Grown under Aerobic and Denitrifying Conditions

Cytochrome c production and release from membranes, was assayed by spectrophotometric analysis of cell fraction suspensions. These assays were complemented with gel electrophoresis analysis of the same, or identically, obtained cultures.

4.1.2.1 Total cytochrome content in aerobic and denitrifying cells

Harvested cells were washed once in 0.01M ammonium acetate buffer, pH 8.5, and resuspended in the same buffer, 1g of cells/5ml Lysozyme (0.05mg/ml) and DNAase (0.01mg/ml) was added, and the solution incubated for 4 hours at 37°C. The resulting lysate was centrifuged at 100.000 g for 1 hour and 15 min. The supernatant was designated the cytoplasmic fraction, and the resuspended pellet (0.01M ammonium acetate, initial volume) as the membrane fraction.

Cytochrome c concentration was found to be negligible in the cytoplasmic fraction of both aerobic, and denitrifying cultured cells.

It was difficult to get consistent spectrophotometric measurements of the cytochrome content of crude membrane fractions. Resuspended membrane pellets were opaque, and dark particulate matter visible. It was therefore decided to dissolve the membrane pellet in detergent.

The membrane pellet was resuspended in 0.05M ammonium acetate, pH 8.5, 2% Triton to give the initial volume. The solution was gently shaken at 4°C for 30 min and then clarified by 5 min spin in a microcentrifuge. The

pellet became considerably smaller, with dark sticky material at the bottom, and then a yellowish layer of looser consistency. The supernatant was clear and dark red. Sometimes, a red translucent layer formed on top of the pellet, which increased with longer spins (later analysis indicated that this might be a b-type cytochrome(s), see section 4.4) but it was easily suspended with gentle suction, using a Pasteur pipette.

Cytochrome c concentration in supernatants of Triton extracted cell membranes, was estimated on the basis of oxidised/reduced difference spectra at A552. An extinction coefficient of $20mM^{-1}cm^{-1}$ was used (horse cytochrome c). The total cytochrome c content was found to be similar in aerobic and denitrifying cultured cells, on the average 25 nmol/g cells. The alpha-peaks of cytochromes c in the extracts, from both aerobic and denitrifying cultures had a maximum absorption at 552 nm, and the alpha/beta ratio was similar. The only noticeable difference was a the slight elevation at 560-570 nm in the anaerobic spectra.

4.1.2.2. The number and sizes of cytochromes

Cells were grown in L-broth medium at high aeration, and anaerobically with nitrate (0.5% KNO3). One study included also cells, grown under low aeration.

Membrane pellets were suspended (1g/4ml) in ammonium acetate buffer, pH 8.5, 2% SDS. A sticky threadlike noncoloured material (DNA/RNA?) was invariably found in the solutions. It did not sediment, even with prolonged centrifugation, but it was possible to decant most of the clear red-brownish supernatant away.

The adequate amount of loading on the gel was determined empirically. 15 µl of solution were found to be adequate which amounts to approximately 1.0-3.0 nmol of cytochromes. 15 µ l of gel loading buffer was added and heated for 3 min at approximately 100°C. The stacking gel method of Laemmli (1970) was used (15% SDS). The samples were electrophorized for 24 hours at constant voltage (3v/cm). In the particular experiment, whose results are shown in Figure 4.1, the gel was washed in methanol (40%) acetic acid (10%) overnight, with shaking, in an attempt to remove noncovalently bound haem. The gel was stained specifically for haem and photographed, then destained and restained for protein with Coomassie blue.

Figure 4.1a shows the gel stained for haem and figure 4.1b shows the gel stained for haem and protein. Figure 4.1c shows a plot of log10 marker molecular mass versus migration distance. The molecular masses of the

cytochromes were calculated from this graph. The bands in the gel are numbered in descending order.

It is evident from figure 4.1a that an abundance of cytochromes is expressed under both aerobic and denitrifying conditions. The main difference between the cells, is the induction in aerobically grown cells of at least two cytochromes, >>68 kD (band 1) and 34 kD. Also noteworthy, is the apparent same concentration per unit membrane weight of cytochromes, common to the aerobically and anaerobically grown cells.

Cells grown under low aeration gave almost identical band pattern to the one from well aerated cells, shown in figure 4.1a. The only difference was an absence of the high molecular weight cytochrome c (band 1).

The doublet bands may be two different electrophoretic forms of the same cytochrome. This has been observed for cytochrome c4 (Pettigrew and Brown, 1988).

4.1.2.3. Attempts to identify c-type cytochromes by gel electrophoresis analysis

Retention of the protohaem by b-type cytochromes in SDS-PAGE gels has been reported (Goodhew *et al.*, 1986), as well as the entrapment of the protohaem by other proteins. Therefore, it is not certain that bands which appear after haem staining, are cytochromes c.

Retention of protohaem by cytochromes b can be reduced or eliminated by a prior haem extraction with organic solvents such as acidified acetone (Goodhew *et al.*, 1986) and ethanol/acetone. Both acid/acetone and ethanol/acetone haem extractions were tried on membrane samples. Samples treated by the former method became insoluble, but could be redissolved to a degree if treated by the latter method. Substantial loading amounts, 45 µl were, however, needed to get visible bands (figure 4.2) compared to 15 µl for untreated samples. As different cytochromes c may have different solubility, it is difficult to interpret the results. It can possibly be maintained, that cytochrome visible bands (even only slightly), after the above treatment, are cytochromes c, whereas an absence does not necessarily mean the contrary. Solubility after the above treatment may be related to molecular weight as well as hydrophilicity. That could explain the relatively high intensity of the two smallest cytochromes, compared to the others. They are probably more resilient to irreversible denaturation than the larger ones, due to small size, low proportional hydrophobicity and compactness.

The three smallest cytochromes are, on the basis of size, more likely to

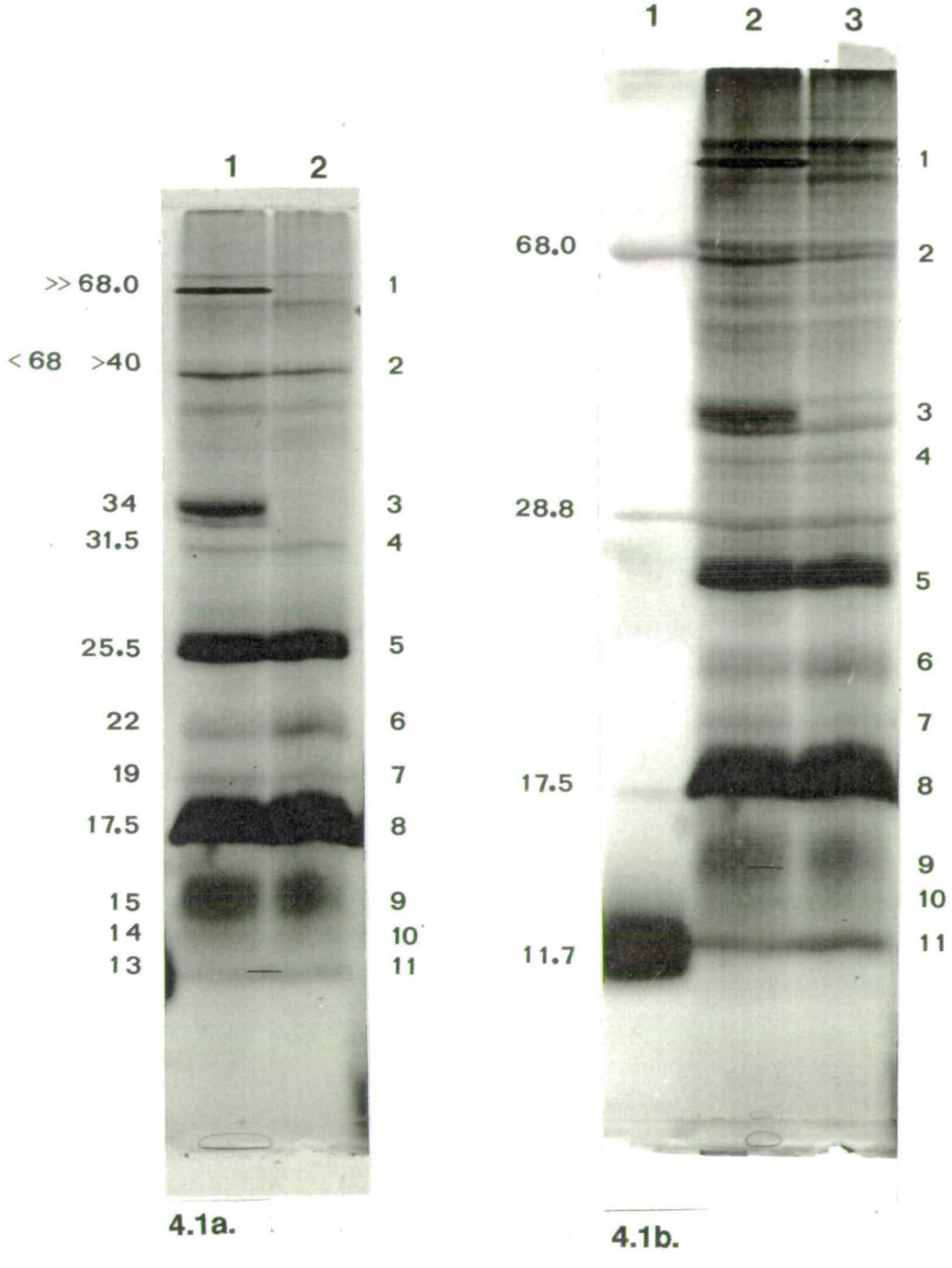


Figure 4.1a. The gel stained for haem only.
The most intense haem stained bands are numbered in descending order from the top of the gel.
Lane 1: The membrane fraction of aerobically grown cells.
Lane 2: The membrane fraction of denitrifying cells.
Figure 4.1b. The gel stained for protein and haem.
Lane 1: Markers; Bovine serum albumin (68000 kD), Penicillinase (28800 kD), Beta lactoglobulin (17500 kD), Cytochromes c (horse, 11700 kD).
Lane 2 : The membrane fraction of aerobically grown cells.
Lane 3: The membrane fraction of denitrifying cells.

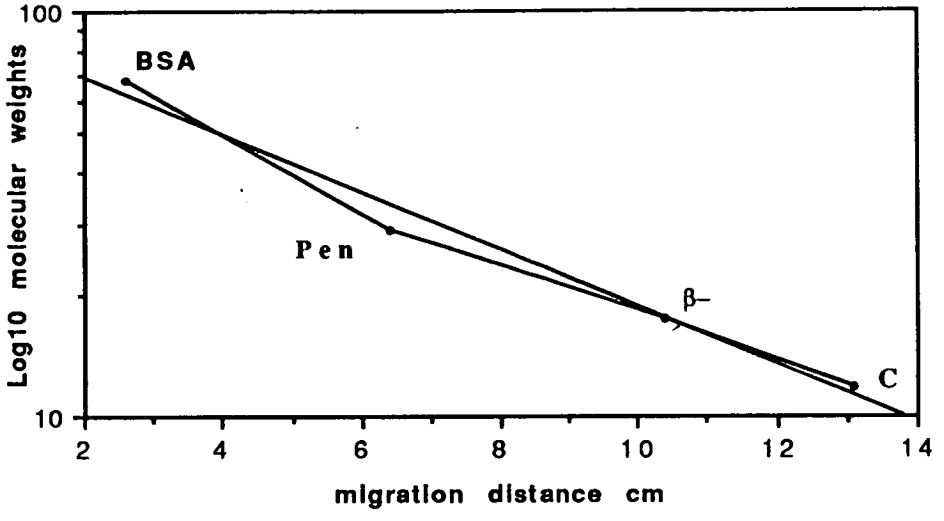


Figure 4.1c. The figure shows a plot of log10 marker molecular mass versus migration distance.
In the graph, the straight region of the graph is extended to cover the range for which a gel of 15% polyacrylamide concentration gives a straight line relationship between molecular mass and migration distance (12000-45000),(Hames, 1990). The molecular masses of the cytochromes were calculated from this graph.
BSA: Bovine serum albumin; **Pen**: Penicillinase; β-: .β-lactoglobulin; C: Cytochrome c (horse).

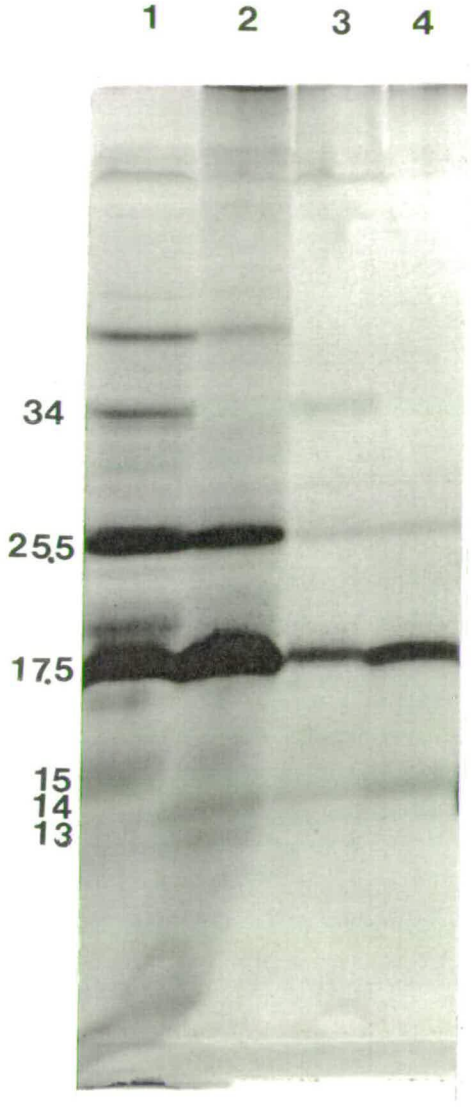


Figure 4.2. SDS-PAGE analyses of membrane cytochromes from *B. azotoformans*, after haem extraction with ethanol acetone.
A portion of membrane fractions of aerobic- and anaerobically grown cells were extracted with ethanol acetone (see methods) and analysed in parallel with untreated membranes, by stacking 15% SDS-PAGE. The differently treated samples were adjusted to the same volume, and 15 µl of untreated membranes and 45 µl of ethanol acetone extracted cells were loaded. After electrophoresis, the gels were fixed in methanol (40%) acetic acid (10%) for 1 hour, and stained for haem with TMBZ in acidic methanol.
Lane 1: Untreated membrane fraction of aerobically grown cells.
Lane 2: Untreated membrane fraction of denitrifying cells.
Lane 3: Ethanol acetone extracted membranes of aerobically grown cells.
Lane 4: Ethanol acetone extracted membranes of denitrifying cells.

be cytochromes c, than either b- or a-type cytochromes.

4.1.2.4. Release of cytochromes from denitrifying cells

The following methods were tried: Acetone extraction, high salt (2.5M NaCl), high alkali (pH 11), and sonication, all of which have been used to release peripheral membrane proteins. Four proteases; trypsin, chymotrypsin, pronase and subtilisin, were also compared. The pellet was resuspended in 0.01M ammonium acetate, pH 8.5, and the solution modified according to the method used. The criteria of solubility was taken to be the retention of proteins in the supernatant, after ultracentrifugation at 100.000 g for 1 hour and 15 min. The concentration of cytochromes c in the supernatant was estimated on the basis of oxidised/reduced difference spectra at 552 nm. An extinction coefficient of 20mM-1cm-1 was used (horse cytochrome c).

No, or negligible release of cytochromes was observed with high salt, high alkali, acetone extraction or sonication. The sonication disrupted the membrane considerably and cytochromes were partially solubilised, but they precipitated by ultracentifugation. The high alkali method, where the desired pH (pH 11) was obtained by adding few drops of concentrated NH3, considerably reduced the size of the pellet, but the haem proteins precipitated upon ultracentrifugation. As many peripheral proteins solubilised under these conditions, and the buffer is highly volatile, this method was considered a convenient early purification step (see also section 4.2.2.1).

The cytochromes appeared to be fully solubilised in 2% Triton x100, but were incompletely solubilised in 2% deoxycholate. In the latter extraction, the smallest cytochromes (band 8 and 11 in figure 4.1a) were not solubilised. They may have been denaturated.

Protease treatment was successful in releasing cytochromes from the membrane fraction of *Bacillus azotoformans* (see section 4.1.2.4.1).

4.1.2.4.1. Comparison of proteases

Two concentrations of proteases were used, 0.01mg/ml and 0.10 mg/ml, and the solutions incubated for 4 hours. The approximate proportional release of total cytochromes was: Subtilisin, 40%; trypsin, 35%; pronase, 15%; and chymotrypsin, less than 5%. As protease treatment may result in too extensive degradation, samples were also analysed by SDS polyacrylamide gel electrophoresis. Samples were electrophorized on a nonstacking discontinuous 4%(3cm)/20%(14cm) SDS gel, pH 8.3 (figure

4.3a). To get some idea of the hydrophilicity of the proteolytic products, native electrophoresis was performed as well, on an otherwise identical gel (figure 4.3b).

Trypsin gave the most acceptable results. Two well demarcated, and seemingly stable hydrophilic bands, were observed. Chymotrypsin did not have any effect. Pronase gave one band and subtilisin one as well, but it furthermore substantially degraded all the cytochromes.

Differences in specificity may explain these results. They can, however, also be interpreted as reflecting the different domain structures of the cytochromes. Rather unexpectedly, two small cytochromes, probably corresponding to bands 10 and 11 in figure 4.1a, migrate in the native gel. This may partly be explained by a small size, but possibly also by a low proportional hydrophobicity. If the haem domains of cytochromes c are approximately the same size, the hydrophobic membrane attachment part will get proportionally smaller with smaller total size of the protein. The two small cytochromes on the SDS-PAGE (bands 10 and 11) appear to be relatively resistant to proteolysis by all proteases. If their membrane attachment-moieties are proportionally smaller than the corresponding parts of the bigger cytochromes c, their haem domains may be in a closer proximity to the membrane and, consequently, the stearically protected from proteolytic attack on the sequence *between the haem domain and the hydrophobic membrane part.

4.1.2.4.2. Comparison of proteolytic release of cytochromes from aerobic and denitrifying cells

Figure 4.4 shows samples of trypsin treated membranes, from both aerobic and denitrifying cultures, that were electrophoresed on a stacking 20% SDS polyacrylamide gel. Two concentrations were used of trypsin, 0.01 mg/ml and 0.10 mg/ml. Samples of untreated membrane preparations were run for comparison. Same amount from each sample preparation (15µl) was loaded and volumes of the untreated samples were adjusted so that membrane weight per ml buffer of each sample, was the same. The figure shows more haem polypeptide bands than were observed in the protease comparison experiment (figure 4.3). The points worth noting are:1) There does not seem to be a marked differences between the proteolytic peptide patterns from aerobic and denitrifying cells.

2) The aerobically induced cytochrome band >>66 kD is apparently not completely digested, and the haem moiety remains membrane bound.

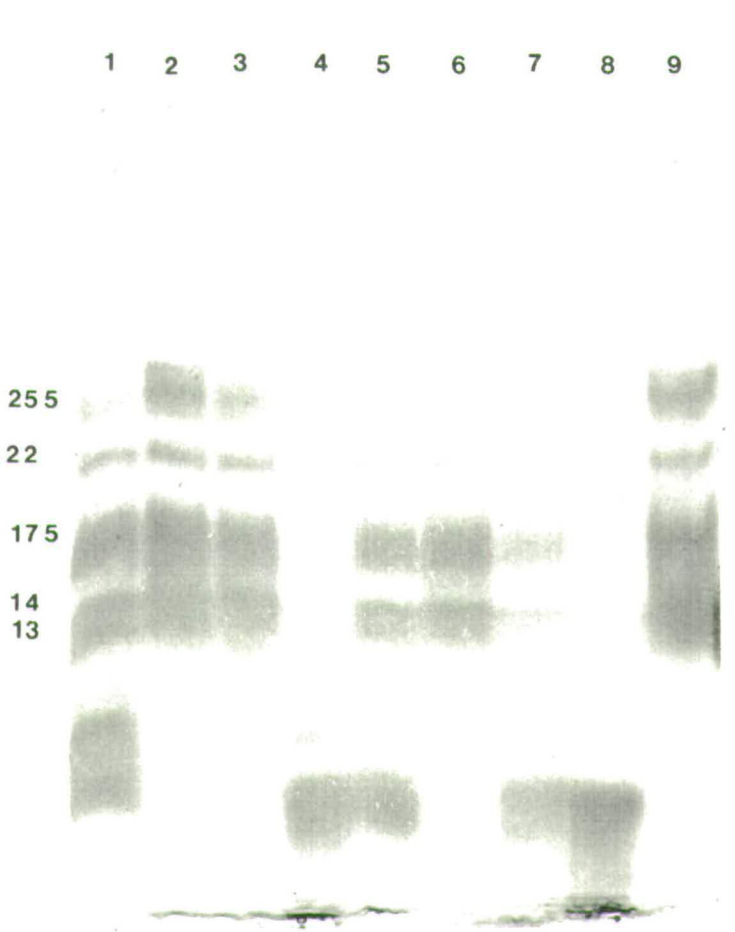


Figure 4.3a. Denaturating gel electrophoresis.

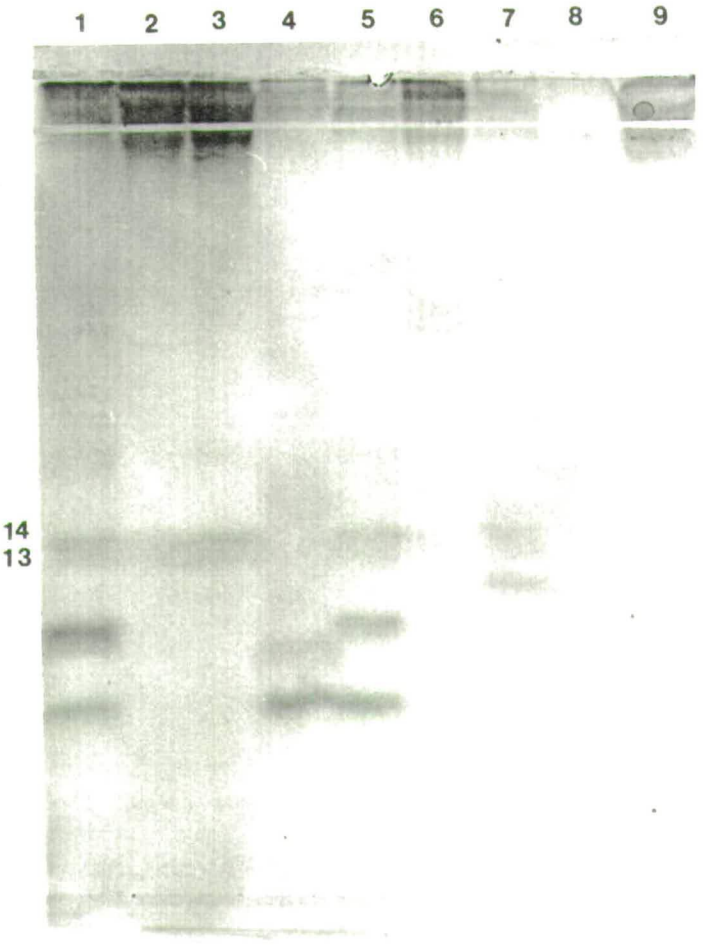


Figure 4.3b. Nondenaturating gel electrophoresis.

Figure 4.3. Proteolytic release of membrane cytochromes from *B. azotoformans* with different proteases.
Portions of membrane fractions of denitrifying cells were treated with trypsin, chymotrypsin, pronase and subtilisin. Two different concentrations, 0.01 and 0.10 mg/ml of proteases, were used. Incubation time was 4 hours. Equal amounts of membranes were loaded in each lane (approximately 1-3 nmols/haem). After electrophoresis, the gels were fixed in methanol (40%) acetic acid (10%) for 1 hour, and then stained for haem with TMBZ in acidic methanol. Samples were electrophorised on a nonstacking discontinuous 4%(3cm)/20%(14cm) SDS gel, 0.025 M Tris-HCl, pH 8.5. Samples were electrophorised on a nondenaturating gel as well, identical to the one described above, but without SDS. Identical samples were loaded from left to right on each gel.

Lane 1: Digestion with trypsin 0.01 mg/ml.
Lane 2: Digestion with chymotrypsin 0.01 mg/ml.
Lane 3: Digestion with pronase 0.01 mg/ml.
Lane 5: Digestion with trypsin 0.10 mg/ml.
Lane 6: Digestion with chymotrypsin 0.10 mg/ml.
Lane 7: Digestion with pronase 0.10 mg/ml.
Lane 9: Untreated membranes.

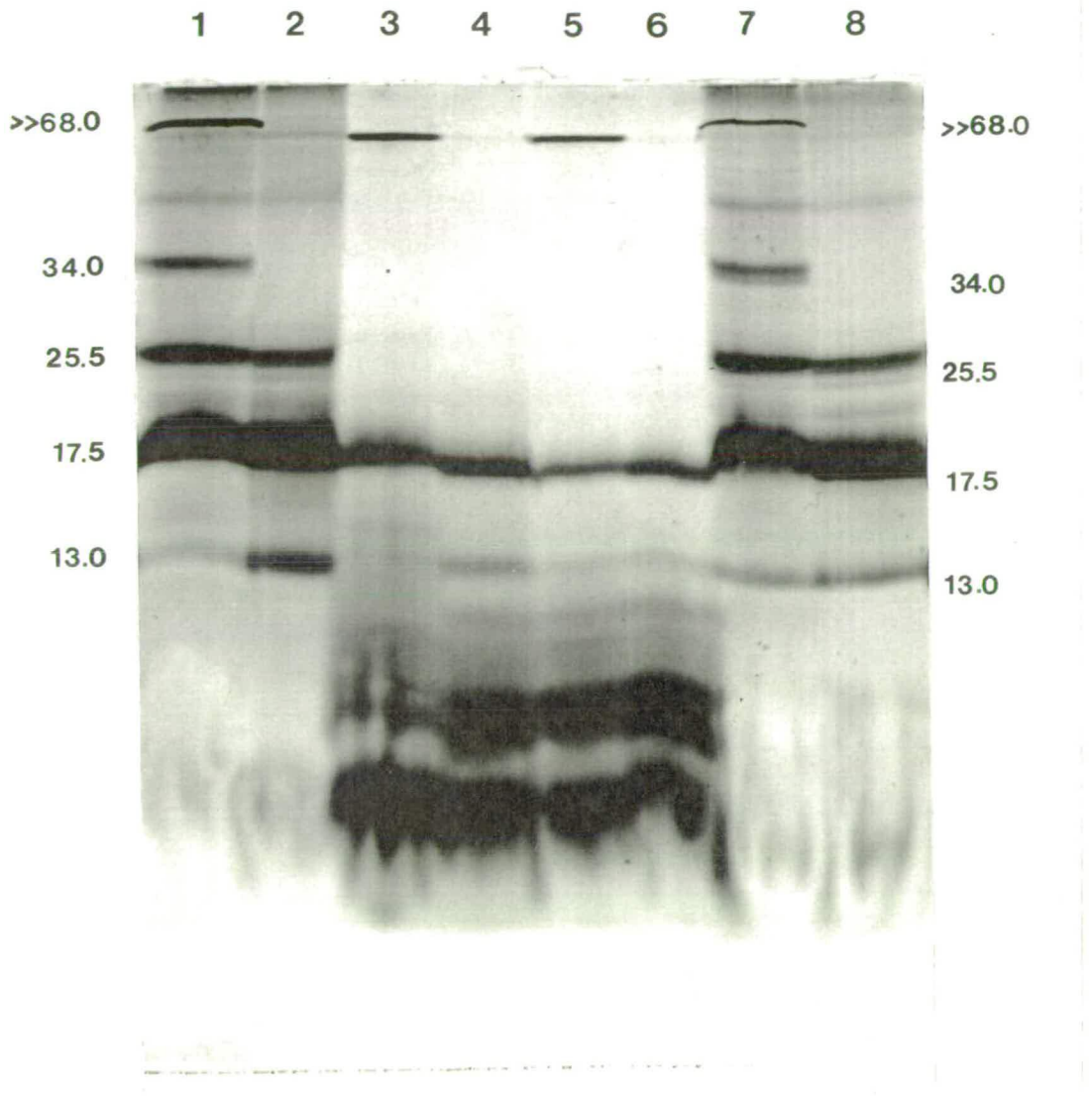


Figure 4.4. SDS-PAGE analyses of trypsin treated membrane cytochromes from aerobically grown, and denitrifying cells of *B. azotoformans*

Portions of membrane fractions of aerobic and denitrifying cells were treated with trypsin. Two different concentrations, 0.01 and 0.10 mg/ml, of the protease were used. Incubation time was 4 hours. Equal amounts of membranes were loaded in each lane. The samples were analysed in parallel with undigested membranes, by stacking 20% SDS-PAGE, using the buffer system of Lemmli (see methods). After electrophoresis, the gels were fixed in methanol (40%) acetic acid (10%) for 1 hour, and then stained for haem with TMBZ in acidic methanol. Samples in lanes 1,3 and 5 and 2, 4 and 6 are from the same culture, respectively. Samples in lanes 7 and 8 are from different cultures.

Lane 1: Untreated membrane, aerobic cells.
Lane 2: Untreated membrane, denitrifying cells.
Lane 3: Aerobic cells, 0.01 mg/ml trypsin.
Lane 4: Denitrifying cells, 0.01 mg/ml trypsin.
Lane 5: Aerobic cells, 0.01 mg/ml trypsin.
Lane 6: Denitrifying cells, 0.01 mg/ml trypsin.
Lane 7: Untreated membrane, aerobic cells.
Lane 8: Untreated membrane, denitrifying cells.

3) Cytochromes, corresponding to bands 8 and 11, and possibly 10, are apparently not extensively digested. However, the bands become somewhat slimmer with increased concentration of protease. It probably reflects an unspecific degradation.

4) The released cytochromes seem to be rather resistant to further proteolysis, as apparently the higher trypsin concentration does not change the peptide pattern and the intensity of the bands to a marked degree.

4.1.3. Choice of Medium for Large Scale Cultures

On the basis of the preliminary studies, the conditions for large scale growth of cells to obtain maximum cytochrome c yields, should preferably be aerobic. However, large scale aerobic cultures were prone to contamination by faster growing opportunistic species and very careful handling, at all stages was necessary. This may not be an unreasonable demand, but the contamination threat always looms when culture conditions are scaled up, because of a greater number of steps and the difficulties in obtaining a complete sterilisation of large fermentors.

The danger of contamination can, apparently, be reduced by growing the strains under denitrifying conditions. It has been observed that denitrifying bacteria utilise, sequentially, the different intermediate nitrogen compounds in the pathway. At an early stage in the culture, the nitrite concentration becomes highly elevated. This substance is very toxic to most bacteria and may therefore hold contaminating species back, to some extent. Large scale cultures were therefore mainly grown under denitrifying conditions.

The media used in the above preliminary small scale studies was Luria-broth (L-broth). This is an expensive medium so only "yeast extract medium" (1% yeast extract and 1% NaCl, pH 7.0, supplemented with $NaNO_3$ (0.5%) for anaerobic cultures), was used. Electrophoretic analysis of the membrane fractions from cells, grown on yeast extract only, showed an identical pattern of cytochromes, to the one obtained from cells grown in the L-broth medium. The cell yield was, however, lower (4-10%).

4.1.4. Summary and Comments

1) An abundance of membrane-bound cytochromes is expressed, both under aerobic and denitrifying conditions, but there is no certainty of which of them are cytochromes c. The conservative assessment, based on acetone/ethanol extraction of the membrane fraction before SDS-PAGE analysis (figure 4.2), was, that there might be up to 5 cytochromes c expressed.

2) Two cytochromes were induced under aerobic conditions, one of which a high molecular weight cytochrome (>>60.000) is repressed under low aeration.

3) The cytochromes, common to both aerobic and denitrifying cells were, apparently, expressed in the same amounts per unit membrane weight and none was conspicuously induced under denitrifying conditions instead two cytochromes were induced under aerobic conditions. This contrasts with the common observations of increased production of cytochromes under denitrifying conditions.

4) Cytochromes c could be released by detergents and proteases, but not by high salt, high alkali or sonication.

5) Of the proteases, trypsin appeared to be best suited for releasing cytochromes c from the membranes

4.2. PURIFICATION OF CYTOCHROMES FROM *B. AZOTOFORMANS*

Figure 4.5 shows a scheme for the purification procedure of cytochromes from *B. azotoformans*, and Table 4.1 shows the purification process for one particular batch of denitrifying cells (batch 3). Table 4.2 shows the release and recovery of the different cytochromes c from different culture batches.

4.2.1 Growth And Harvesting of Cells

Bacillus azotoformans was stored in glycerol at -20°C. It was streaked onto Luria broth plates (L-plates) and incubated at 37°C for 24-48 hours. These plates were kept at room temperature. The bacteria seem to be more viable at this temperature, being retrievable for more than a month, compared to less than a week when kept at 4°C.

10 ml cultures were grown from a single colony in L-broth. These were used for inoculating 1-2l cultures which were used as inoculum, 2.0-2.5%, for the large scale cultures. The cultures were examined for contaminating bacteria, by streaking a loopful onto L-agar plates. The large scale cultures were either grown in L-broth medium or in "yeast medium"

4.2.1.1. Anaerobic cultures

Anaerobic cultures were grown in 40 litre batches overnight, for approximately 16 hours. Four such batches were grown at a time and processed together. The cultures evolved great quantities of gas, and frothing was pronounced. At the end of the culture period, some lysis was always observed, but no contamination problems occurred.

4.2.1.2. Aerobic cultures.

Aerobic cultures were grown in 50 litre batches. Even with careful handling of the inoculum, the large batches were, more often than not, contaminated with opportunistic species, though they rarely did overtake the culture completely. Sometimes, contamination became evident as soon as the cells were harvested, because of the unfamiliar colour and texture of the cell precipitate, but usually the contamination was only observed when the cells were washed. Then, an unexpected layering was observed after the centrifugation stage. In some cases, the apparent contamination was confirmed by microscopic analysis. The most plausible explanation is that faster growing bacteria were not eliminated when the fermentor was sterilised.

4.2.1.3. The harvesting of cells

Cells were harvested by centrifugation in a continuous flow centrifuge. The yields from anaerobic cultures were 3.5-5.0 g/l, with the greatest yield from the culture grown on L-broth (table 4.2, batch 1). The cell yield from the only uncontaminated aerobic culture obtained, was 5.6 g/l (table 4.2, batch 5) which was grown in "yeast medium". The cell precipitates had a distinct red tinge, which was a degree deeper in the anaerobic cultures. The cells were washed twice with 0.01 M ammonium acetate buffer, pH 8.5, and processed immediately.

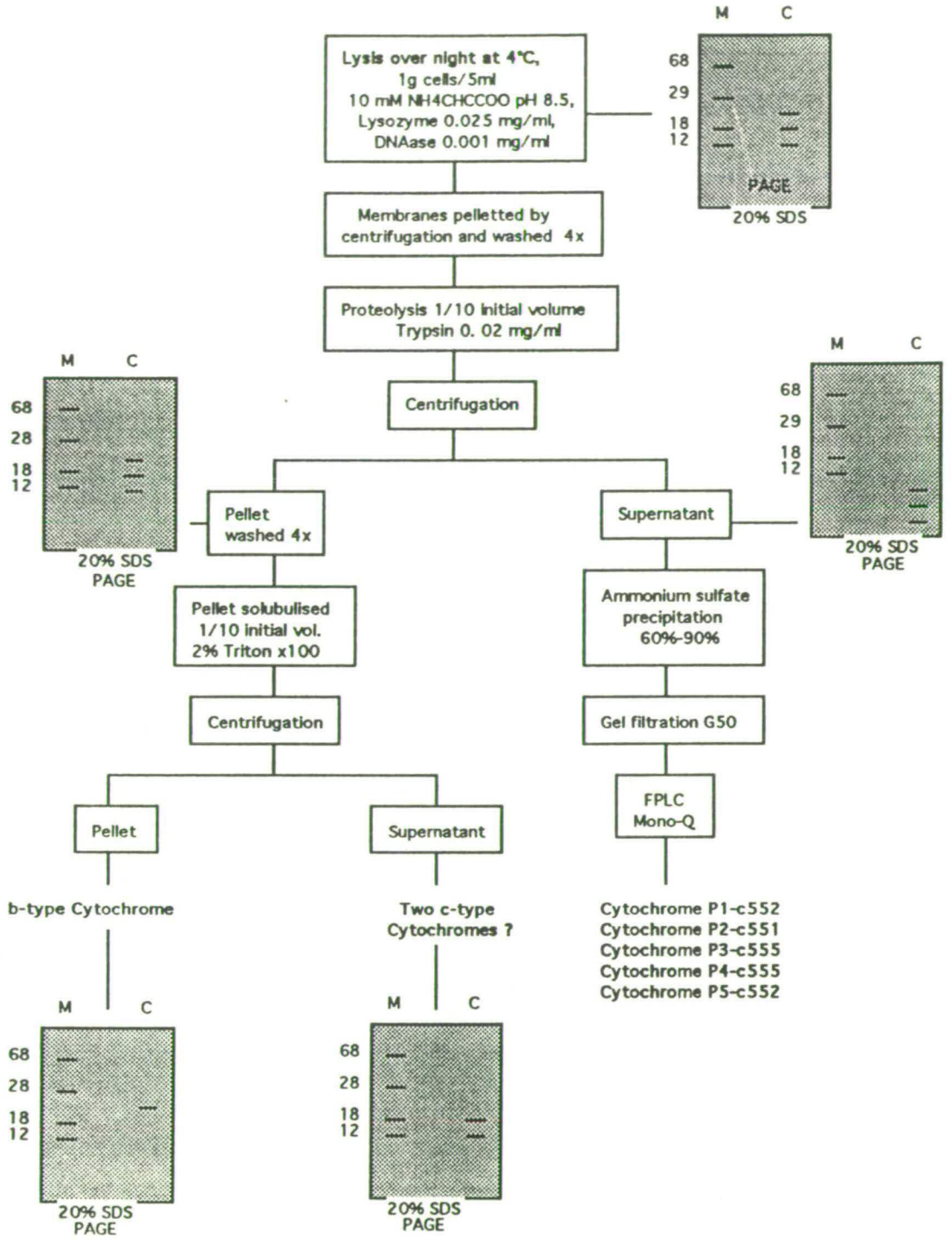


Figure 4.5. A scheme of purification procedure of cytochromes c from *B. azotoformans*.

Table 4.1. Purification of cytochromes c from 560g cell (wet weight) of *Bacillus azotoformans* anaerobically grown with nitrate.

Fraction	Total protein mg	Total cyt. c µmol	nmol cyt./ mg protein	Yield [1] total cytochromes c %	Recovery[2] released cytochromes c %	Purification[3]
Membranes		13.44[4]		100		
Trypsin digest: supernatant	4450	5.81	1.3	43	100	1
Am. sulfate precipitation	141	4.37	40.9		75	31
Gel filtration	35	3.51	100.3		60	77
FPLC eluate Mono-Q		1.49[5]	—			

Cytochrome c concentration was estimated by using the molar extinction coefficient, 20 mM-1 cm-1, at 552 nm, reduced minus oxidised. Protein was estimated by the method micro Lowry (see methods).

[1] Recovery is relative to the total amount of cytochrome c in the membrane fraction, on the basis of measurements of Triton extracted membrane of a denitrifying culture of *B. azotoformans*.

[2] Trypsin releasable cytochromes.

[3] Purification is relative to the nmol cyt./mg protein of trypsin released cytochromes c in the supernatant. It is assumed that cytochromes not released by trypsin are of a different type than the ones released

[4] Measured in Triton extract of cell membranes of *Bacillus azotoformans* from small scale cultures.

[5] The total amount of cytochromes c in the FPLC eluate, the sum for the all the collected cytochrome c peaks.

Table 4.2. Release of cytochromes c and yields from different large scale cultures of *Bacillus azotoformans*.

Batch	culture[1] method	Yield cells g	cytochromes released µmol	cytochromes nmol/g cells	Yield total cyt. µmol (%)	P1-c552 yield µmol (%)[3]	P2-c551 yield µmol (%)[3]	P3&P4-c555 yield µmol (%)[3]	P5-c552 yield µmol (%)[3]	other cyt.c µmol (%)[3]	ratio[4] P3&P4/P5
1	*anaerobic*	789	6.29	7.8	2.94 (46)	1.49 (51)	0.27 (9)	0.85 (29)	0.33 (11)	-	0.60
2	*anaerobic*	662	5.01	7.6	1.92 (38)	0.96 (50)	0.14 (7)	0.70 (36)	0.12 (6)	-	0.70
3	*anaerobic*	561	5.81	10.4	1.49 (26)	0.87 (58)	0.03 (2)	0.47 (32)	0.01 (1)	0.12[5] (8)	0.55
4	*anaerobic*	559	5.40	9.7	1.45 (27)	0.75 (52)	0.17 (12)	0.41 (28)	n.c.[6]	0.12[7] (8)	0.53
5	*aerobic*	280	2.93	10.5	0.65 (22)	0.41 (63)	n.c	0.24[5] (37)	n.c	-	0.50

[1] Anaerobic cultures were grown with NO3-
[2] Trypsin released cytochromes c.
[3] Percentage of total trypsin released cytochromes in supernatant.
[4] The ratio of the two major cytochromes c isaolated, P1-c552 and P3&P4-c555.
[5] These cytochromes c were only minor peaks of the eluate, with alpha absorption maxima of 555 nm.
[6] n.c. means not collected
[7] This is a cytochrome designated P2-c551b, it was between the P2-c551 and P3-c555 with alpha maximum of 551 nm and electrophoretic properties similar to P2-c551.

4.2.2. Purification Procedure

4.2.2.1. Isolation of membranes

The packed bacteria were suspended (1g/10ml) in 0.01 M ammonium acetate buffer, pH 8.5. Lysozyme (0.025 mg/ml) and DNAase (0.001mg/ml) were added, and the mixture shaken for 4 hours, at 37°C, or overnight at 4°C. The lysate was centrifuged for 2 hours in a GS3 rotor at 8500 g and the supernatant discarded.

The pellet was washed 2-4 times in 0.01 M ammonium acetate, pH 8.5, and each time the supernatant was carefully removed by water suction. The washing of the first two batches of denitrifying cells was done with ammonium acetate, pH 10.5. A considerable reduction in pellet size was observed, but ultra-centrifugation at 100.000 g (60 ml tubes) was required to fully precipitate the red coloured haem proteins. It was found to be more convenient to use a GS3 rotor (500 ml tubes), but as it has a maximum rotor speed of only 8500 g, the high alkali washing step was omitted in the last three batches (batches 3, 4 and 5, see table 4.2).

4.2.2.2. Proteolytic release of cytochromes c

The membrane pellet was resuspended (1g/ml) in 0.01 M ammonium acetate buffer pH 8.5. Trypsin was added, to give a concentration of 0.02 mg/ml. The mixture was incubated for 4-6 hours at 37°C, with constant shaking. The proteolysed membrane solution was centrifuged for 2 hours at 15000 g in a GSA rotor and the supernatant collected by water suction. The total yield of releasable cytochromes c was 7.6-10.4 nmol/g in the denitrifying cultures, and 10.5 nmol/g from the only aerobic culture. The release was, at most, 43% of the total cytochrome c content (24 nmol/g cells) of the cells, as measured in 2% Triton x100 membrane-extracts of small whole cell cultures (see section 4.1.2.1).

The amounts of released cytochromes were less, per unit weight cells, from the membrane preparations washed in high alkali (table 4.2, batches 1 and 2), possibly, because the membranes were disrupted to a greater extent so that a small part of the cytochromes was washed away.

4.2.2.3. Ammonium sulfate precipitation

Ammonium sulfate was slowly added to the supernatant from the proteolysis stage, to give a 60% saturation. The solution was stirred for 30 min at 4°C, and the resultant suspension centrifuged at 10000 g for 15 min. The great bulk of the proteins precipitated. The supernatant was set aside.

The sediment was resuspended by adding water, and brought to 60% saturation again and recentrifuged. The supernatants were then combined, and the precipitate discarded. By bringing the ammonium sulfate up to 90% saturation, the cytochromes c were precipitated. This sediment was greatly enriched in cytochromes and had a distinct red colour. The purification was variable from batch to batch, at best around 40 fold, but most of the time lower, only 20-30 fold. The concentration of cytochromes in the supernatant, after the 90% precipitation, was negligible. Recovery of the total trypsin released cytochromes c was typically around 70% (table 4.2).

4.2.2.4. Gel filtration

The cytochrome c precipitate was suspended in the smallest volume possible, in 0.05 M ammonium acetate buffer, pH 5.1, and put on top of a 1m x 16mm column, packed with Sephadex 50. The elution profile consisted of an early large peak of non cytochrome proteins with a closely trailing smaller peak of a mixture of cytochrome c polypeptides. A shoulder was often observed, just before the main cytochrome peak. Figure 4.6a shows a typical gel filtration elution profile of ammonium sulfate precipitated cytochromes c and figure 4.6b shows the cytochrome c polypeptide composition of the shoulder, and the main cytochrome c peak. Cytochrome fractions were freeze-dried and redissolved in the smallest possible volume.

4.2.2.5. Fast protein liquid chromatography

The cytochrome c mixture from the gel filtration was separated into individual components by Fast Protein Liquid Chromatography (Pharmacia) The cytochromes c did not bind to the cationic Mono-S column, but bound firmly to the anionic Mono-Q mono column. A gradient system was used, consisting of two buffers, one of Tris-HCl, pH 8.5, connected to pump B, and the other of Tris-HCl, pH 8.5, 1 M NaCl, connected to pump A. A gradient elution was developed (which gave a slow linear gradient of 0-30%), over 30 min, and then a steeper gradient of 30-100%, over 20 min. A flow rate of 1ml/min was used. Initially, very diluted samples were run, but concentration and volume were gradually increased, until resolution of the closest peaks became effected. This loading amount was then used throughout, for the remaining sample. The recovery of total cytochromes, after the last purification step (combined collected cytochrome peaks) was greater from membranes

washed in high alkali (pH 11), or 30-40%, (table 4.2, batches 1 and 2), than from membranes which were not, or approximately 25% (table 4.2, batches 3, 4, and 5).

Figures 4.7 and 4.8 show the FPLC elution profiles of cytochromes from aerobic and denitrifying cells, respectively, and the alpha absorption maxima of the cytochrome c in the corresponding fractions. The cytochromes which were collected are designated with a P and a rising number in the order which they were eluted, and their alpha maximum absorption wavelength: P1-c552, P2-c551, P3-c555, P4-c555 and P5-c552, in the order of elution.

The elution profile for cytochromes c, obtained from aerobic and anaerobic cultures was identical and the same cytochromes were released. This was confirmed by SDS-PAGE analysis. This is surprising, as SDS-PAGE analysis of untreated membranes from aerobic and anaerobic cells reveals that one cytochrome, 34 kD, appears to be induced in aerobic cultures and this cytochrome c is digested by trypsin (fig.4.4). This cytochrome may be a b-type cytochrome or if it is a cytochrome c it was either fully degraded or was lost in the purification process. It may have precipitated below the 60% ammonium sulfate saturation step.

The cytochromes were fully oxidised after the gel filtration step, and the FPLC elution profile varied little from batch to batch. The elution profile did, however, change, if the cytochrome c sample was reduced with sodium dithionite before the FPLC. Then, the number of peaks increased. This can be seen in figure 4.9, where the P1-c552 and P4-c555 peaks both split up into two peaks, after reduction with sodium dithionite. For each cytochrome, one peak may represent the oxidised form, the other one the same cytochrome c reduced.

The collected cytochrome c peaks were desalted into 10 mM ammonium acetate buffer on G25 Sephadex and the preparations freeze-dried overnight.

The yield of individual cytochromes varied from batch to batch (table 4.2), but certain trends were observed. The yield of P1-c552 was always greatest, about 50% of the total cytochromes obtained. The other major cytochrome c, c555, which was collected in two peaks, P3-c555 and P4-c555, amounted at most to 40% of the total cytochromes. The ratios of these cytochromes (P3&4-c555/P1-c552) were between 0.50-0.60 from four batches, and as high as 0.70 from one batch (table 4.2).

The proportion of the two minor cytochromes, c P2-c551 and P5-c552, was small. It was highest 20% for P2-c551 from one particular batch

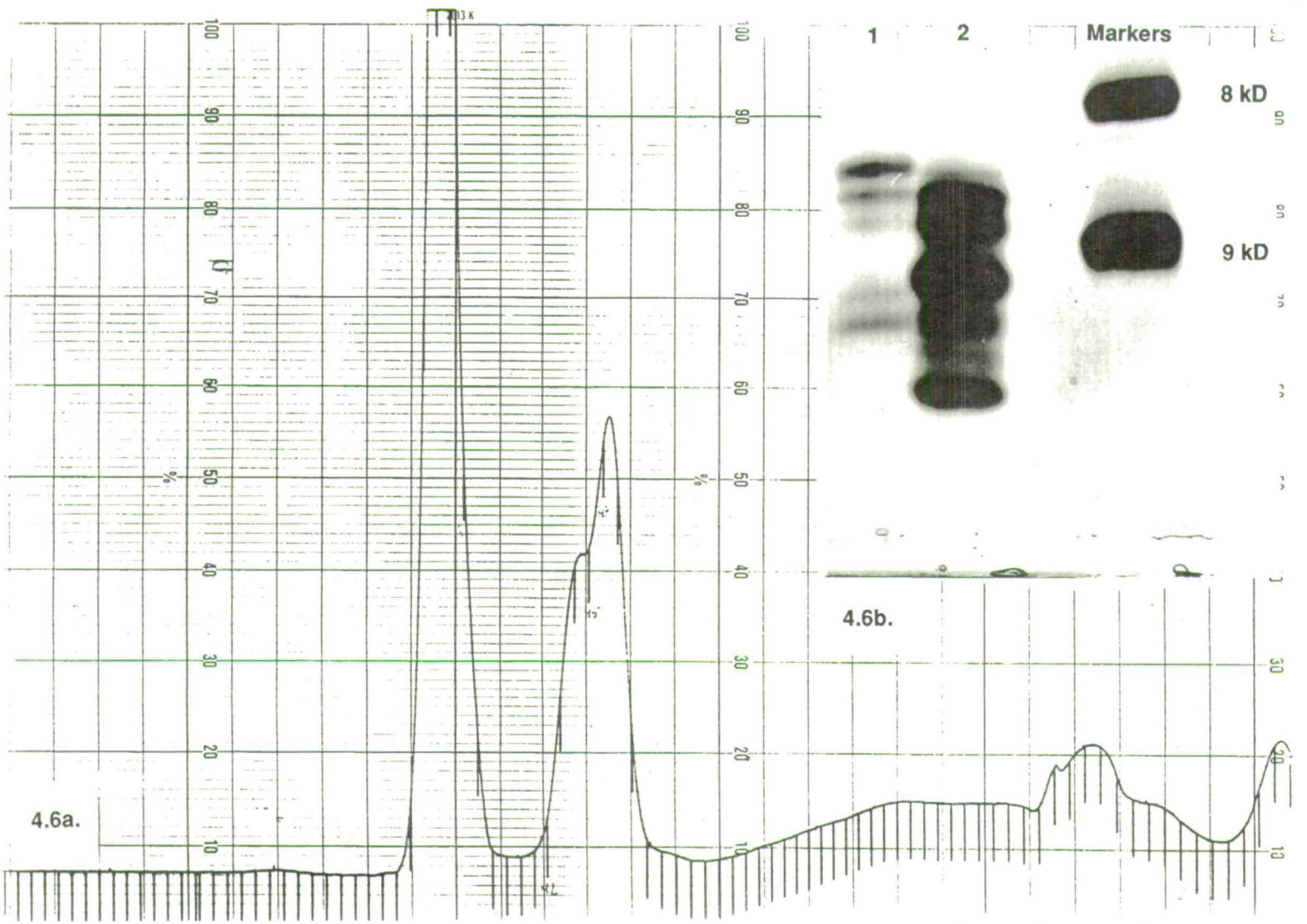


Figure 4.6. Gel filtration of the 60-90% ammonium sulphate fraction of the supernatant from the trypsin treated membranes.

The cytochrome c, 60-90% ammonium sulfate precipitate,was suspended in the smallest volume possible, in 0.05 M ammonium acetate buffer, pH 5.1, and applied to 1 m x 2 cm column, packed with Sephadex 50.

Figure 4.6a. The gel filtration elution profile.

The fractions indicated, contained cytochrome c and were collected.

Figure 4.6b. 20% SDS-PAGE analysis of samples from the shoulder, and the main cytochrome peak.

A stacking gel was used, using the buffer system of Laemmli (see methods). The gel was stained for haem.

Lane 1: Sample from the shoulder of the cytochrome c peak.

Lane 2: Sample from the main cytochrome c peak.

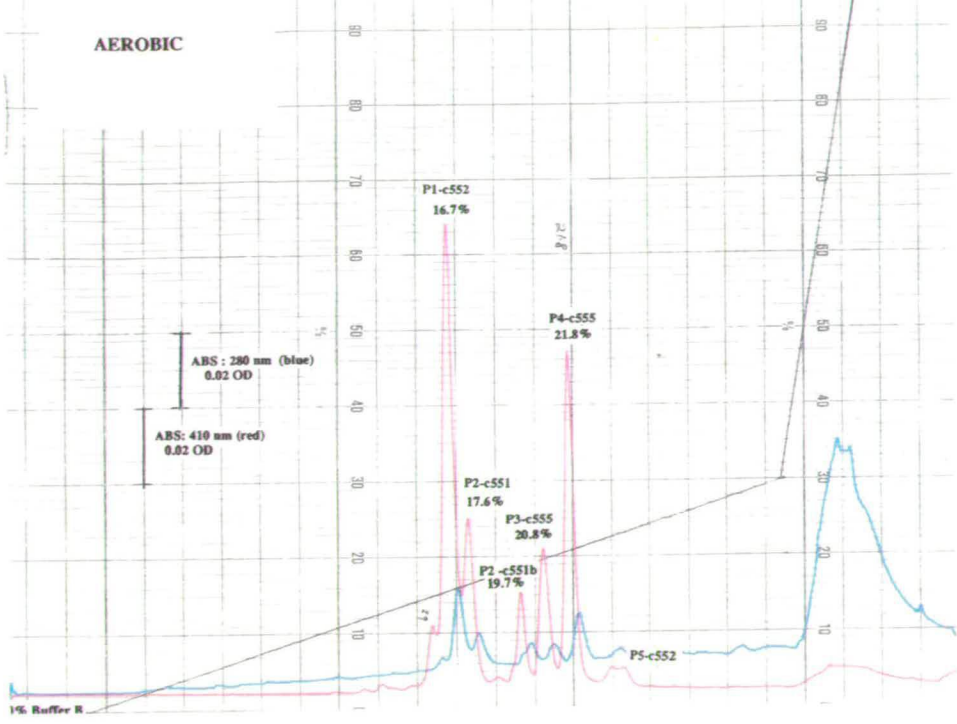


Figure 4.7 Continuous gradient FPLC elution from Mono-Q of cytochromes c from aerobically grown *B. azotoformans*.
The supernatant, after 60-90% ammonium sulfate preparation and gel Sephadex 50 filtration, was applied on a Mono-Q HR 10/10 column. *Flow rate:* 1 ml/min. *Buffer A:* Tris-HCl, pH 8.5. *Buffer B:* Tris-HCl, pH 8.5, 1M NaCl. *Gradient:* 0% B for 5 min., 0-30% B for 30 min., 30-100% B for 20 min.

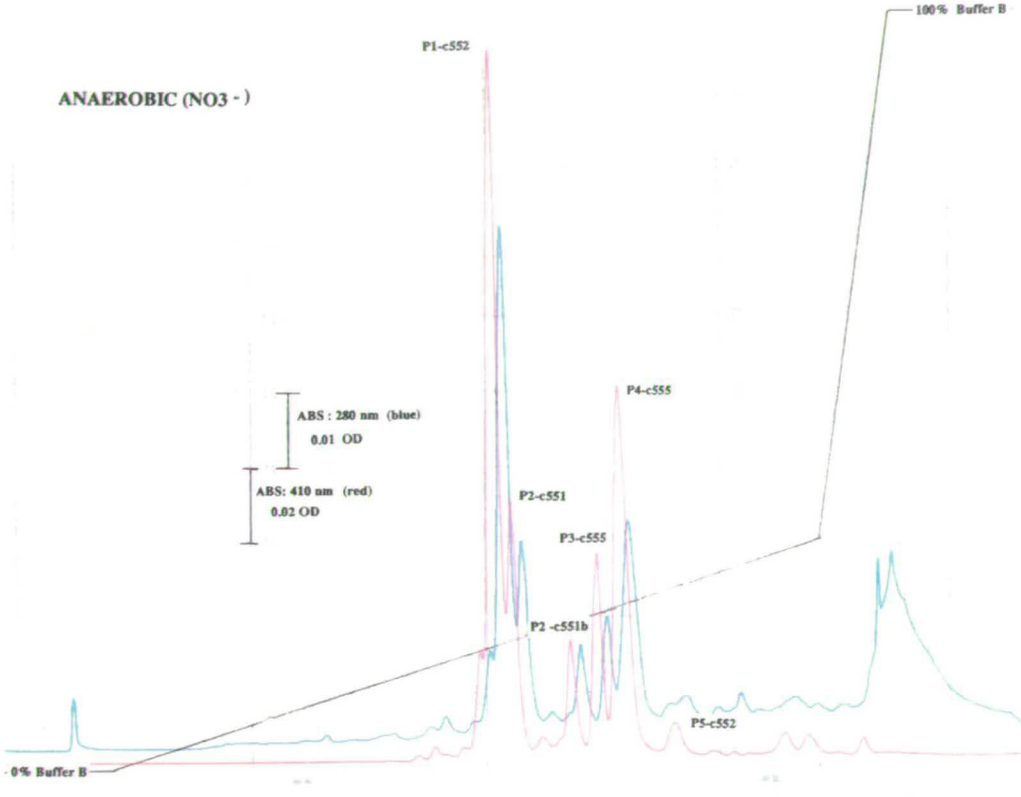


Figure 4.8. Continuous gradient FPLC elution from Mono-Q of cytochromes c from denitrifying culture of *B. azotoformans*.
The supernatant, after 60-90% ammonium sulfate preparation and gel Sephadex 50 filtration, was applied on a Mono-Q HR 10/10 column. *Flow rate:* 1 ml/min. *Buffer A:* Tris -HCl, pH 8.5. *Buffer B:* Tris-HCl, pH 8.5, 1M NaCl. *Gradient:* 0% B for 5 min., 0-30% B for 30 min., 30-100% B for 20 min.

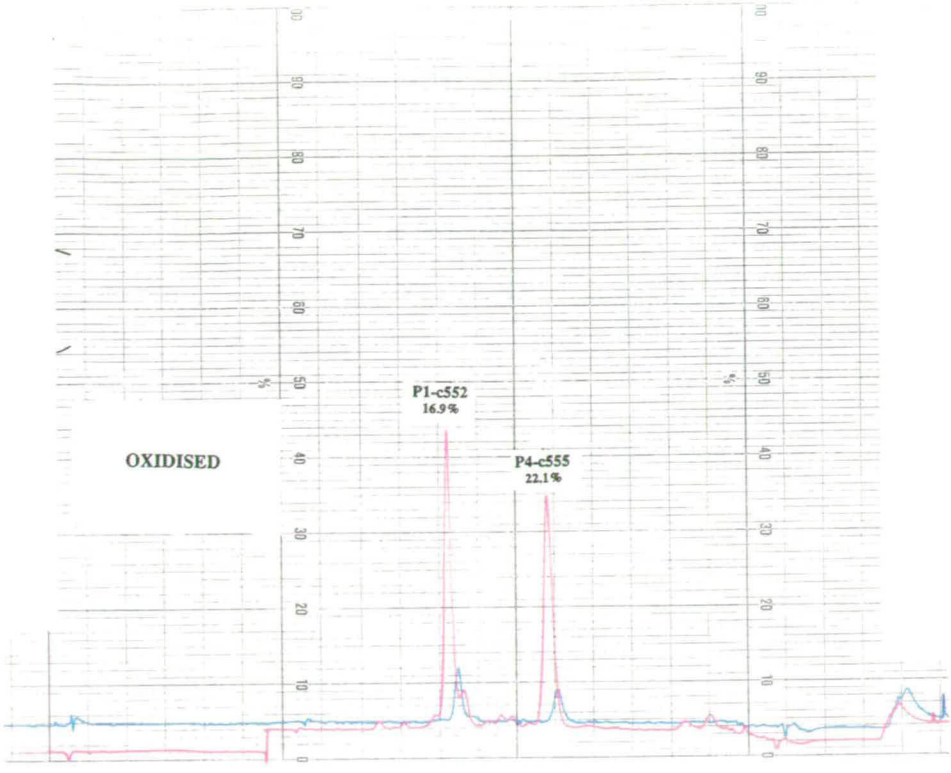


Figure 4.9. Mono-Q FPLC elution profiles of oxidized and reduced mixtures of cytochromes from *B. azotoformans*.
The sample was oxidized with potassium ferricyanide or reduced with sodium dithionite before application on Mono-Q HR 10/10 column. *Flow rate*: 1 ml/min. *Buffer A:* Tris -HCl pH 8.5. *Buffer B:* Tris-HCl, pH 8.5, 1M NaCl. *Gradient::* 0% B for 5 min., 0-30% B for 30 min., 30-100% B for 20 min.

Figure 4.9a. Continuous gradient FPLC elution from Mono-Q of an oxidized mixture of cytochromes P1-c552 and P4-c555.

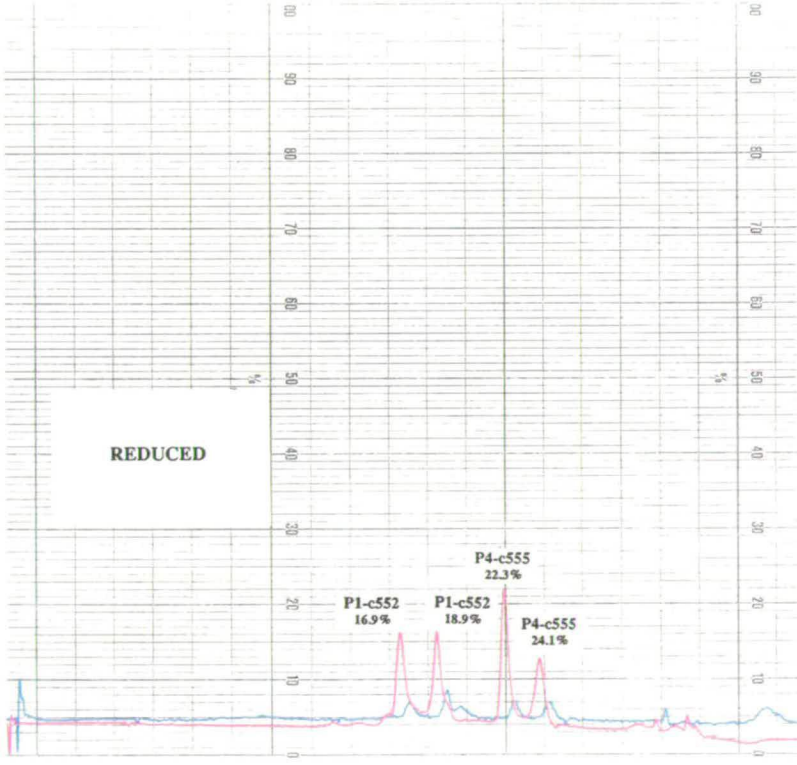


Figure 4.9b. Continuous gradient FPLC elution from Mono-Q of a reduced mixture of cytochromes P1-c552 and P4-c555.

(batch 3), collected in two peaks (P2-c551 and P2b-c551)(batch 4). The P2b-c551 peak is shown in figures 4.7 and 4.8 and lies between the P2-c551 and P3-c555. According to SDS-PAGE analysis this peak consists mainly of one component, the smallest one of the three bands, which was found in peak P2-c551.

The yield was most 11 %, for P5-c552, of the total cytochromes c yielded. It appears, that preparations made from high alkali washed membranes, were enriched in this minor component cytochrome.

The minor peaks were not collected from all batches, because of too small yield.

4.3. CHARACTERISATION OF CYTOCHROMES C, PURIFIED FROM *BACILLUS AZOTOFORMANS*

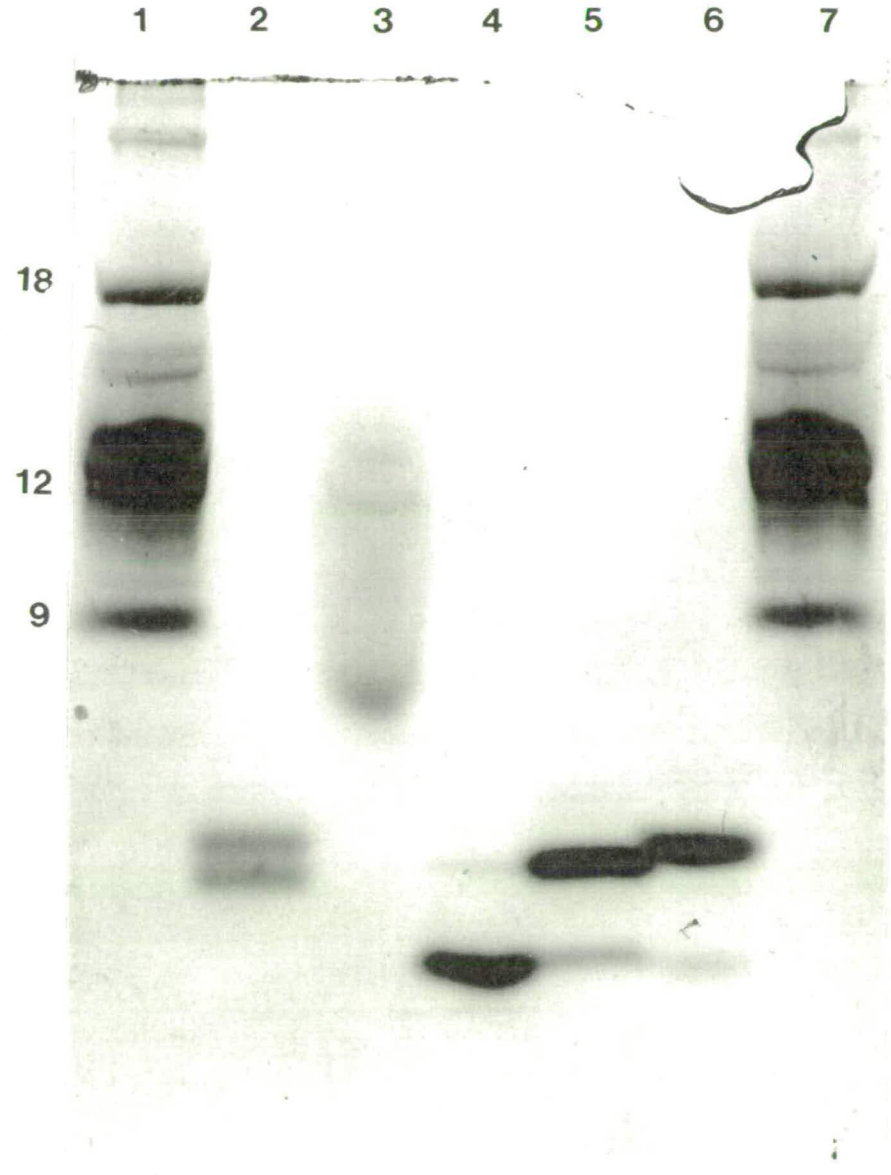
4.3.1. SDS-PAGE Analyses of Purified Cytochromes

The purity of FPLC cytochrome c isolates, was evaluated with regard to contamination with different cytochromes c, and non-cytochrome proteins.

The different FPLC preparations were analysed electrophoretically on 20% SDS gels. They were stained separately for haem and protein (figure 4.10a&b). A minor non-cytochrome component was observed in the P5-C552 fraction. It was estimated to be less than 10% of the total protein.

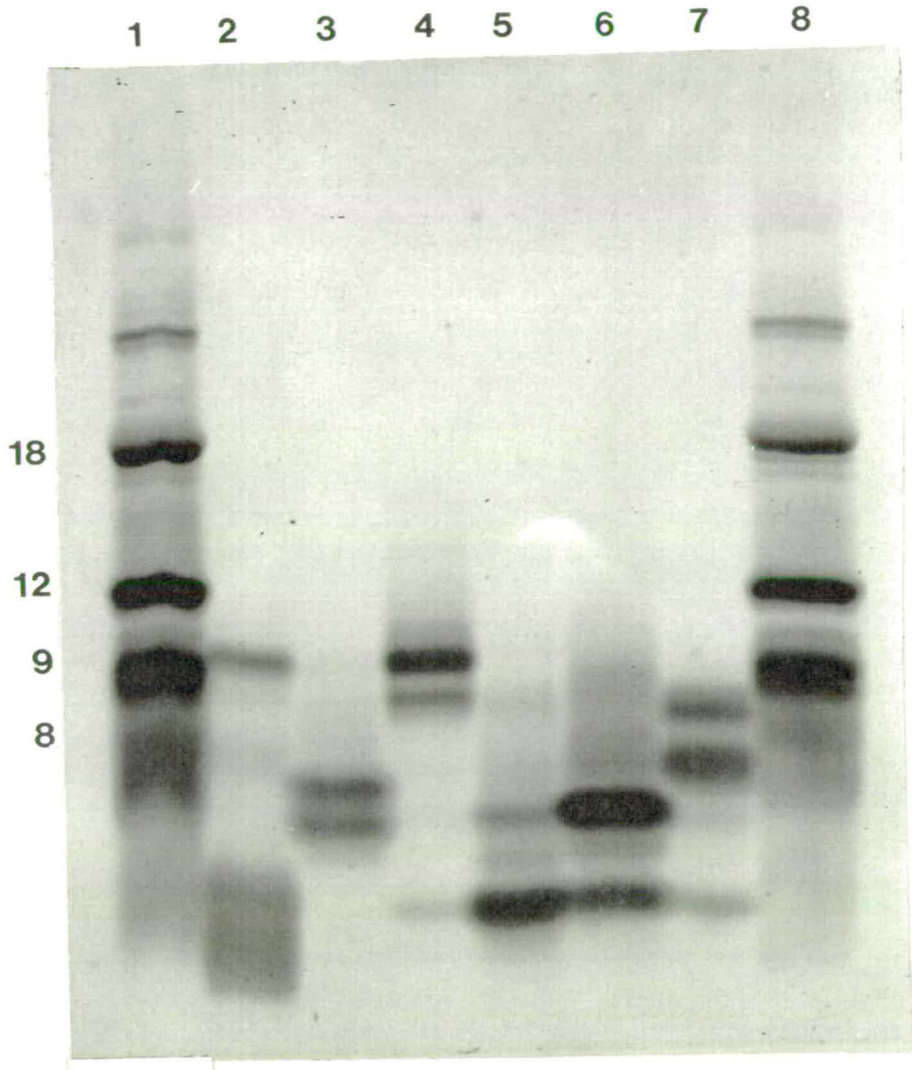
All of the collected FPLC fractions gave more than one haem staining band on the SDS-PAGE analysis. The banding pattern varied somewhat from batch to batch, but was characteristic for each fraction. The fractions showed little cross-contamination. A greater number of bands became visible with increasing loadings. The haem stain is very sensitive, it is enzymatic, and minor cytochrome impurities are amplified relative to the major bands that become saturated relatively quickly (Goodhew *et al.* 1986). These contaminating cytochromes were not observed when the same gels were stained for protein by Coomassie blue (figure 4.10b). The redox potential titration experiments (see below) suggest, that the cytochrome c bands, visualised on the gels from each fraction, represent fragments of the same precursor molecule.

The P1-c552 showed 2-3 bands, of which one was the most intense . This cytochrome was not very susceptible to the Coomassie stain, but the receptiveness to the stain increased with repeated, and prolonged washing



4.10b.

Figure 410, continuation.
Figure 4.10b. Protein staining of trypsin released cytochrome c from *B. azotoformans*, separated by FPLC.
Lane 1 and Lane 7: Markers; *Pseudomonas stutseri* c4 (18000 kD), c5 (9000 kD), and horse cytochrome c (11700 kD).
Lane 2: P2-c551, approximately 2 nmols.
Lane 3: P5-c552, approximately 2 nmols. This cytochrome precipitated upon storage in the freezer, with resultant streaking during electrophoresis.
Lane 4: P3-c555, approximately 2 nmols.
Lane 5: P4-c555, approximately 2 nmols.
Lane 6: P1-c552, approximately 2 nmols.



4.10a.

Figure 4.10. SDS-PAGE analyses of FPLC fractionated cytochromes c.
Samples of cytochrome c fractions from the FPLC separation were analysed by 20% SDS-PAGE. Samples, of the same preparation, were also analysed for amino acid composition (table 4.3). After electrophoresis, the gels were fixed in methanol (40%) acetic acid (10%) for 1 hour, and then either stained for haem with TMBZ in acidic methanol or for protein with Coomassie Blue (see methods).

Figure 4.10a. Haem staining of trypsin released cytochromes c from *B. azotoformans*, separated by FPLC.
Lane 1 and Lane 8 : Markers; *Pseudomonas stutseri* c4 (18000 kD), c5 (9000 kD), c551 (8200 kD), horse cytochrome c (11700 kD).
Lane 2: P5-c552, approximately 200 pmols. This sample was further treated with 0.01 mg/ml trypsin for 4 hours at 37°C.
Lane 3: P2-c551, approximately 200 pmols.
Lane 4: P5-c552, approximately 200 pmols.
Lane 5 : P4-c555, approximately 200 pmols.
Lane 6 : P3-c555, approximately 200 pmols.
Lane 7: P1-c552, approximately 200 pmols.

with a fixing solution (10% Glacial acetic acid and 40% methanol). Only one band showed up on the Coomassie stained gel.

The P2-c551 showed from two to three bands, a doublet with bands of same intensity, and sometimes a larger band appeared of varying intensity, from some batches.

Peak P2-c551b had the same alpha absorption maxima as the P2-c551 and very similar electrophoretic characteristics on SDS-PAGE gel. It was therefore concluded, that they derived from the same protein.

P3-C555 and P4-c555 were cross-contaminated, but were later shown to be derived from the same cytochrome.

P5-C552 showed up as three bands. The largest one was the most intense one.

4.3.2. Peptide Mapping

P1-c552 and P5-c552, as well as P3-c555 and P4-c555, had the same alpha absorption maxima. In order to determine if these distinct cytochrome c fractions, separated by FPLC, were different cytochrome "species" but not different fragments of the same protein, they were subjected to peptide mapping by HVPE pH 6.5. The results are shown in figure 4.11. It is apparent, that the cytochromes P3-c555 and P4-c555 are different fragments of the same protein, whereas the maps of P1-c552 and P5-c552 are too dissimilar to be of the same protein.

4.3.3. Amino Acid Composition

Cytochromes from the different FPLC fractions, P1-c552, P2-c551, P3-c555, P4-c555 and P2-c552, were subjected to amino acid analysis as described in methods . The results are shown in table 4.3.

The amino acid composition of the P3-c555 and P4-c555 cytochromes is very similar, and the smaller one can be derived from the other. This was supported by comparative peptide mapping.

The amino acid compositions of P1-c552, P2-c551, P3-c555 & P4-c555 and P552-c552 are sufficiently different to conclude, that they are different cytochromes c. This was supported by the comparative peptide mapping for P1-c552, P5-c552, P3-c555 and P4-c555.

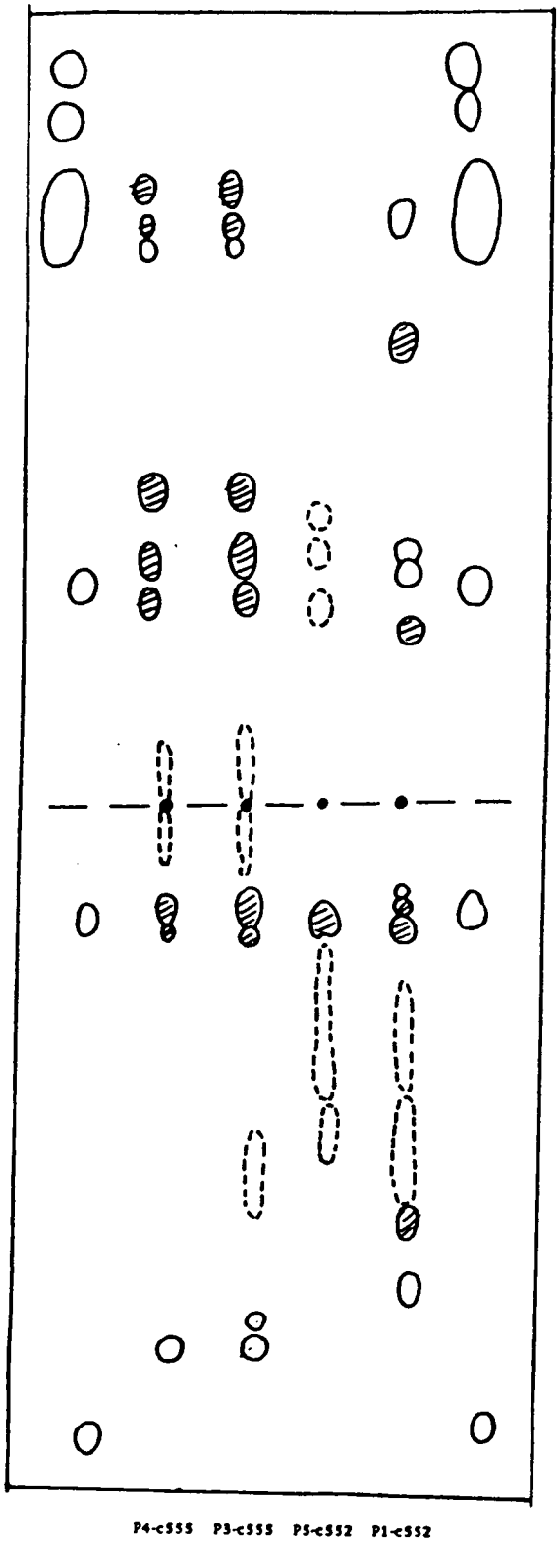


Figure 4.11. HVPE peptide mapping of cytochromes c released by trypsin from membranes of *B. azotoformans.*
50 nmol of each cytochrome, P1-c552, P3-c555, P4-c555 and P5-c552 were digested with pepsin in 5% formic acid, pH 2.5, for 4 hours at 37°C. The samples were dried down and dissolved in 30 µl of 0.1M NH_3 and 20 nmol spotted (10 µl at a time) onto a Whatman 3MM paper. The peptides were separated in parallel by HVPE at pH 6.5. The paper was stained for peptides with ninhydrin.

Table 4.3. Amino acid composition of cytochromes isolated from *Bacillus azotoformans*.

Cytochrome P1-a552.

amino acid	nmols	number
Lys	4.22	4
His	0.96	1
Arg	1.00	1
Asx	9.85	10
Thr	1.51	1-2
Ser	5,99	6
Glx	12.2)	12
Pro	4.47	4-5
Gly	10.52	10
Ala	18.81	19
Val	5.19	5
Met	2.83	3
Ile	3.27	3
Leu	4.60	5
Tyr	0.17	0
Phe	0.26	0
Cys	1.98	2
Trp		1
Total		89

Cytochrome P5-a552.

amino acid	nmol	number
Lys	9.08	9
His	2.37	2
Arg	1.00	1
Asx	12.04	12
Thr	13.68	14
Ser	8.94	9
Glx	24.18	24
Pro	5.56	6
Gly	13.88	14
Ala	13.4	13
Val	8.04	8
Met	0.88	1
Ile	6.76	7
Leu	8.10	8
Tyr	2.51	2-3
Phe	3.96	4
Cys	1.78	2
Trp	N.D.	
		137

Cytochrome P3-a555

amino acid	nmols	number
Lys	7.14	7
His	2.08	2
Arg	1.00	1
Asx	9.19	9
Thr	8.27	8
Ser	6.41	6-7
Glx	15.65	16
Pro	5.86	6
Gly	13.68	14
Ala	15.00	15
Val	3.44	3-4
Met	1.81	2
Ile	4.85	5
Leu	10.15	10
Tyr	1.29	1
Phe	3.72	4
Cys	N.D	(2)
Trp	N.D	
Total		113

Cytochrome P4-a555

amino acid	mmols	number
Lys	5.93	6
His	2.00	2
Arg	1.00	1
Asx	9.80	10
Thr	7.74	8
Ser	6.22	6
Glx	12.02	12
Pro	5.71	6
Gly	12.44	12
Ala	14.43	14
Val	2.64	3
Met	1.91	2
Ile	3.70	4
Leu	9.84	10
Tyr	0.99	1
Phe	3.71	4
Cys	2.38	2
Trp	N.D	
Total		103

Ctochrome P2-a551

amino acid	nmol	number
Lys	7.00	7
His	2.08	2
Arg	0.52	1
Asx	12.00	12
Thr	3.82	4
Ser	2.42	2-3
Glx	9.98	10
Pro	3.53	4
Gly	13.46	14
Ala	18.52	19
Val	3.45	4
Met	0.83	1
Ile	1.57	2
Leu	6.89	7
Tyr		0
Phe	0.96	1
Cys	2.01	2
Trp*		1
Total		93

Amino acid analysis of peptides and proteins were performed on a Beckman 120C amino acid analyser. The amount of amino acid (nmol) per petide or protein were estimated on the basis of the peak area of a detected amino acid by reference to standards of known amount.

Shown below are the deduced numbers of putative haem or iron binding residues, and aromatic amino acids.

P1-c552:	2C 1H 2M 1W
P2-c551:	2C 1H 2M 1W 1 F 1Y
P3&P4-c555:	2C 2H 2M 2Y 1F (trp not estimated)
P5c-552:	2C 1H 2M 2Y (trp not estimated).

According to the SDS-PAGE analyses, every FPLC collected cytochrome fraction consisted of different proteolytic fragments. In each fraction, they are believed to be fragments of the same precursor cytochrome c, on the basis of spectral analyses and redox potential measurements (see section 4.3.7). The presence, in the fractions, of different fragments obviously effects the composition analyses to a greater or lesser extent.

Other factors which can distort the measurements are:

1) Contamination by different proteins. It is not absolutely certain that the bands on the proteolytic cytochrome c fragments that co-purify together, are from the same cytochrome c precursor. A minor non-haem protein contamination was also found in the P5-c552 fraction, although it was estimated to be less than 10%.

2) Errors in calculating the area of the amino acid peaks. Sometimes, the peaks were not well separated and baselines not horizontal.

3) The presence of incompletely hydrolysed bonds such as ile-ile.

4) Contamination by free amino acids. Serines and glycines are commonly observed contaminants.

5) Breakdown of amino acids (tyrosine and serine anatryptophan).

6) Modification of amino acids (glutamines, aspargines and tryptophan).

4.3.4. Sizes of the isolated cytochrome c

The sizes, or molecular weights of the cytochromes, isolated in this study, cannot be defined as definite "properties", as the cytochromes are apparently proteolytic fragments of larger structures. Also, more than one fragment of the same co-purified protein, appear on SDS gels as bands of varying intensities. This apparent controversy between the FPLC elution profiles and the SDS-PAGE analyses indicates, that trypsin co-purified with the cytochromes c, and acted during the consequent purification steps or in the SDS gels.

According to the SDS-PAGE analysis, the isolated cytochrome c polypeptides are all smaller than horse cytochrome c, which is 11.7 kD. Unfortunately, markers of molecular weight less than 8.0 kD were not available, so it was not possible to make accurate size estimation by SDS-PAGE analysis. The most prominent cytochrome bands can, however, be put into the following descending order: P5-c552 > marker 9kD > P1-c552 > marker 8kD > P2-c551 > P3-c555 >> P4-c555.

Molecular weights are calculated from amino acid compositions, in descending order, as following: P5-c552, 13.7 kD; P3-c555, 11.3 kD; P4-c555, 10.3; P2-c551, 93 kD, and P1-c552, 8.9 kD. According to these estimates, the molecular weight of P5-c552 is greater than that of the horse cytochrome c marker, whereas it is the other way around by SDS-PAGE analysis. Also, P3-c555 and P4-c555 are much larger, measured by amino acid analysis, than by SDS-PAGE analysis, and are larger than P1-c552. These size estimates should therefore be taken with a fair measure of scepticism. On the other hand, the molecular weight of P1-c552, calculated directly from the sequence, is approximately 8.2 kD (see Chapter five). The estimate, from the amino acid analysis, of 8.9 kD is therefore reasonably accurate.

The Laemmli (1970) system is not particularly well suited for estimating sizes of polypeptides smaller than 12 kD (Hames, 1990) and factors that may contribute to anomalous behaviour on SDS gels, are incomplete unfolding of the protein and inefficient sieving of small molecules. Anomalous behaviour of cytochromes c on SDS gels, where different estimates of sizes are obtained for the apoprotein and the holoprotein, has been reported (Pettigrew and Brown, 1988). It must be emphasised, that the amino acid composition estimates are derived from fragment mixtures, and interpretation of the results of amino acid composition measurements are, to a degree, subjective. Many amino acids are destroyed, to some extent, during hydrolysis, and some peptide bonds are more resistant to hydrolysis than others. The extinction coefficients may also be different from that of the horse cytochrome c, which was used for calculation of molar concentrations.

4.3.5. N-terminal analysis

A single N-terminal amino acid, was found for all the FPLC peaks collected. Glycine was found for P1-c552, and glutamine or glutamate for all the others; P2-c551, P3-c555, P4-c555 and P5-c552.

4.3.6. Spectra

The absorption spectra of the cytochromes were determined as described in methods. The spectra are shown in figures 4.12. The extinction ratios are listed in table 4.4.

The cytochrome solutions were fully oxidised after the last purification step and were reduced, by sodium dithionite.

The spectra of the P1-c552 and P5-c552 have the same alpha absorption maxima, but different beta and gamma maxima. Their extinction ratios are almost identical, P2-c551 is similar to them, but P3&P4-c555 are distinctly different, although they fall into the general mould of low-spin cytochrome c spectra. With the possible exception of the P3&P4-c555, which have spectra similar to those found for class IC cytochromes c from purple bacteria, the isolated cytochromes have no distinct features which can be associated with spectra of already defined cytochrome c subclasses.

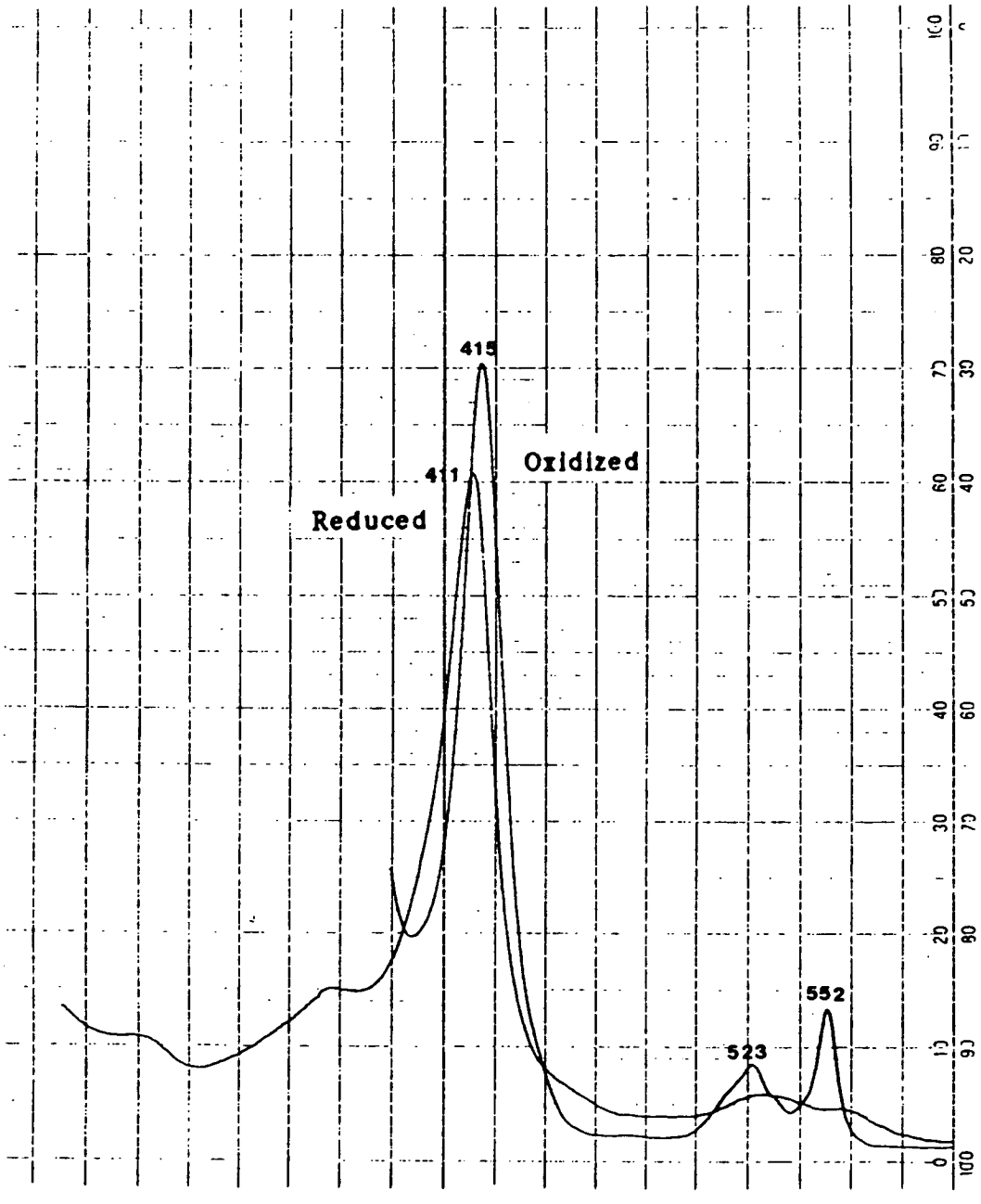
The P1-c552, P2-c551 and P5-c552 have a near symmetrical alpha peak. The P3&P4-c555 cytochromes have a redshifted split-alpha peak, so far, found only in a limited number of cytochromes of class IC. They also have the highest Soret/alpha extinction ratio.

Absorption at 280 nm is unusually low in all four cytochromes, which implies a low or no tryptophan content. This is reflected in the 280/alpha peak ratio which is around 0.7 for all the cytochromes.

4.3.7. Redox potentials

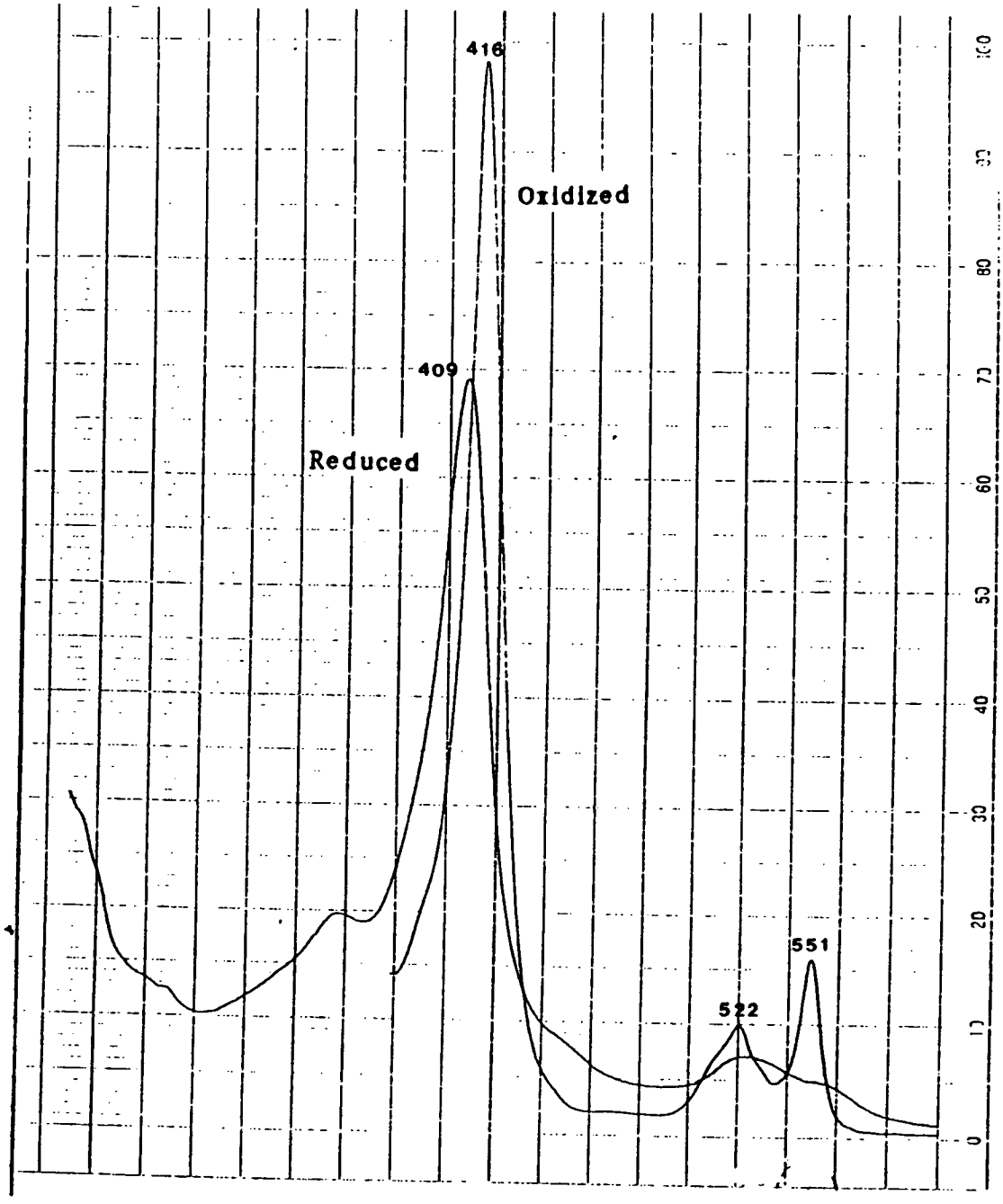
The midpoint redox potentials of the cytochromes were determined for P1-c552, P2-c551, P3&4-c555 and P5 -c552 (figure 4.13 and table 4.4). The redox potentials are in the lower range, normally found for small soluble cytochromes from Gram negative species. The point worth noting is, that the redox potential of split alpha cytochrome P3&P4-c555 is considerably higher than the one reported of the split alpha c554 from *B. subtilis,* or 124mV vs -80mV, respectively. The latter potential, if correct, is unusually low for a c-type cytochrome and lower values are only found among c3 cytochromes from *Desulfovibrio* (approximately -200mV), (Pettigrew and Moore,1987).

It should be noted, that the redox potential titration curves for the different cytochrome c polypeptide fractions, all appear to be of single species cytochrome c. The midpoint redox potentials are comparatively low. They are, in ascending order; 93mV for P2-c551, 122mV for P1-c552, 124mV for P3&4-c555, and 188mV for P5-c552. The slope of the titration

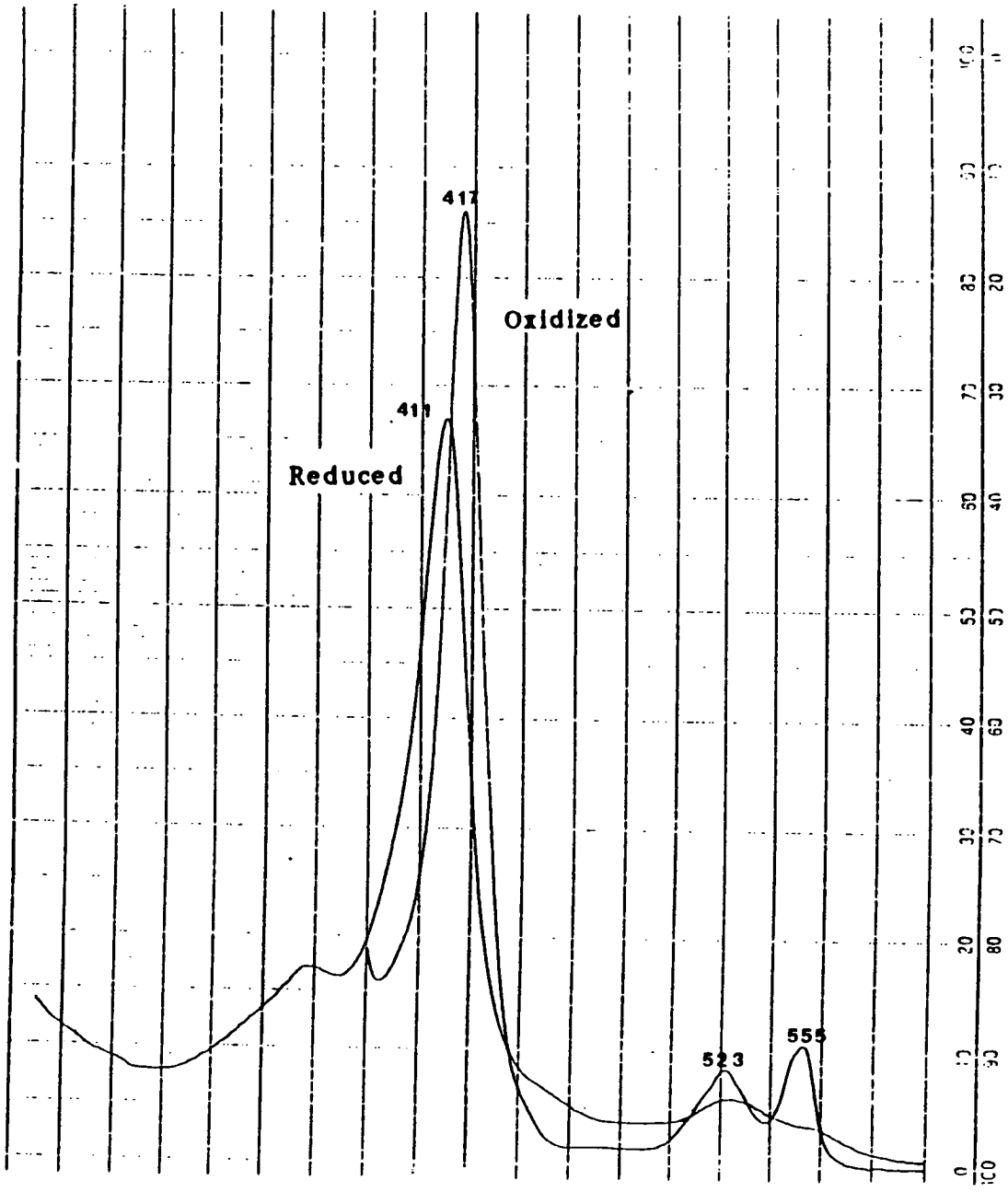


4.12a. Cytochrome P1-c552.

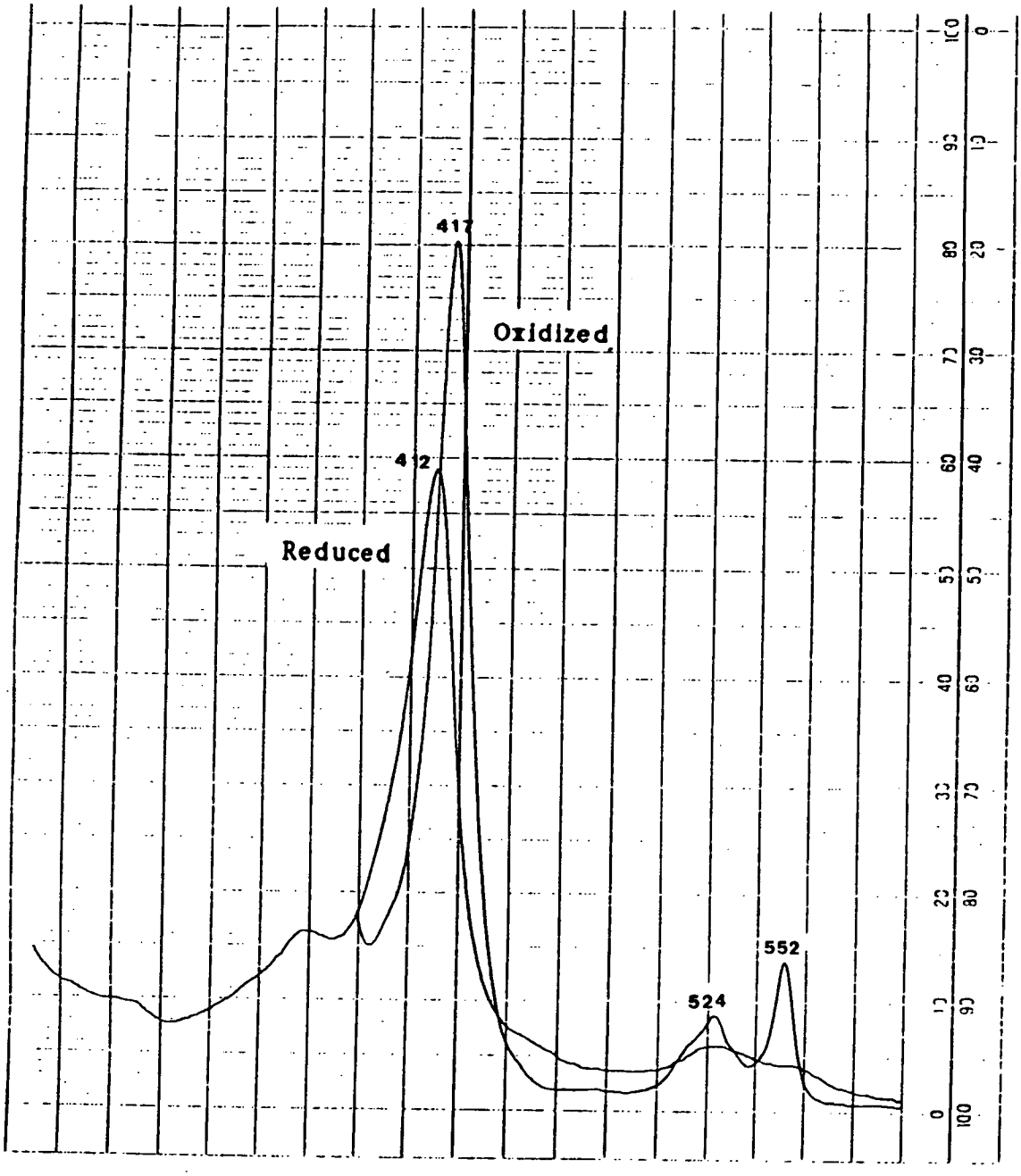
Figure 4.12. UV-visible spectra of the trypsin released membrane bound cytochromes c from *B. azotoformans*.
In each case, the ferri- and ferrocytochrome spectra were obtained by adding potassium ferricyanide and sodium dithionite, respectively, to solutions of the cytochrome in 0.1 M phosphate, pH 7. Absorption maxima and extinction ratios are listed in table 4.4.
Figure 4.12a. Cytochrome P1-c552.
Figure 4.12b. Cytochrome P2-c551.
Figure 4.12c. Cytochrome P4-c555.
Figure 4.12d. Cytochrome P5-c552.



4.12b. Cytochrome P2-c551.

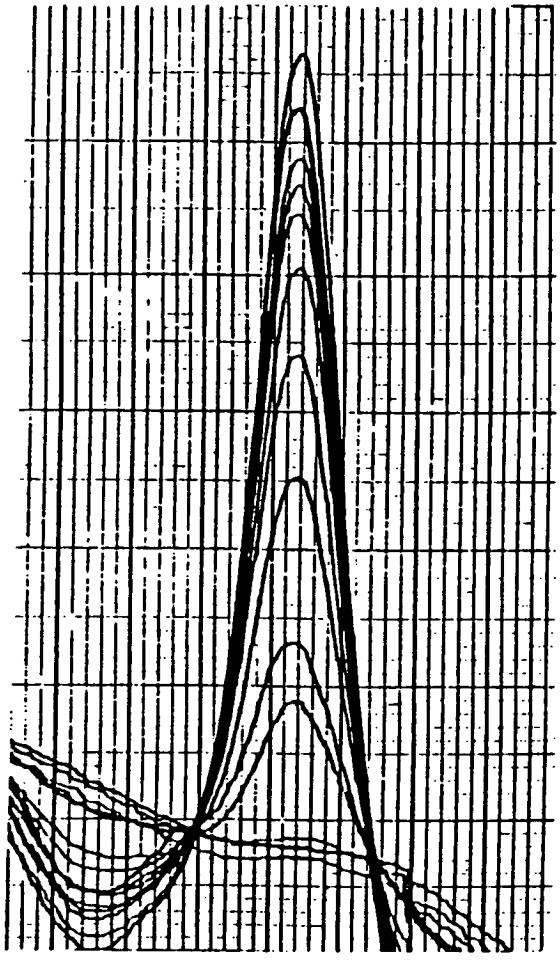


4.12c. Cytochrome P4-c555.



4.12d. Cytochrome P5-c552.

Figure 4.13. Redox titrations of cytochromes c from *B. azotoformans*. Cytochrome spectra were recorded in a stirred anaerobic cuvette, containing 20 mM phenazine methosulfate, diaminodurol and ferric ammonium sulfate, and 10 mM EDTA. Each spectrum corresponds to ambient potential and is used to calculate log cyt_{ox}/cyt_{red}. Experimental points are from oxidative and reductive titrations. The midpoint redox potentials are listed in table 4.4.



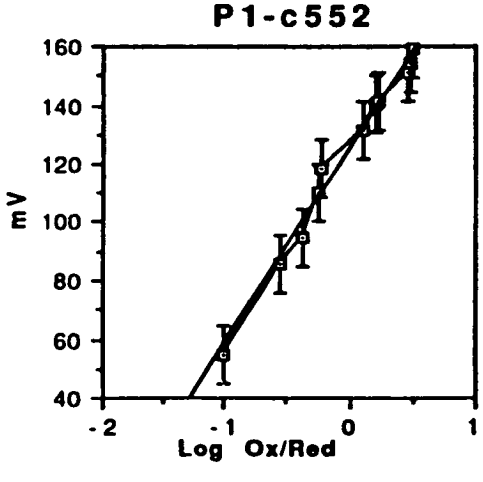
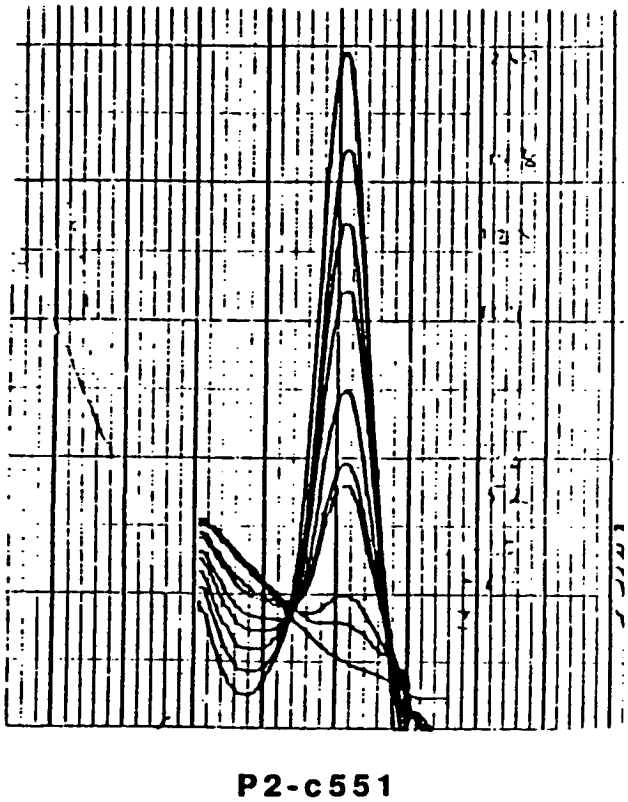


Figure 4.13a. Cytochrome P1-c552. Both oxidative and reductive titration spectra are shown. Slope 66



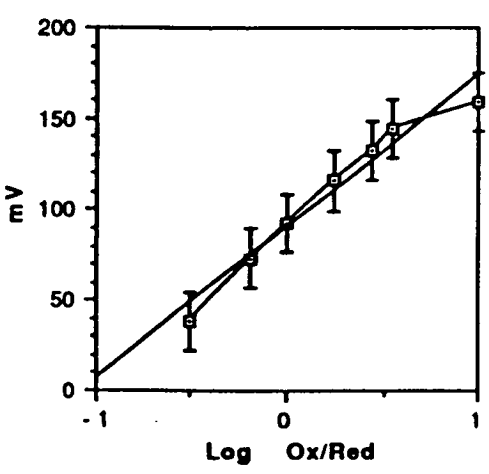


Figure 4.13b. Cytochrome P2-c551. Only the oxidative titration spectra are shown. Slope 75

Figure 4.13. Redox titrations of cytochromes c from *B. azotoformans*.

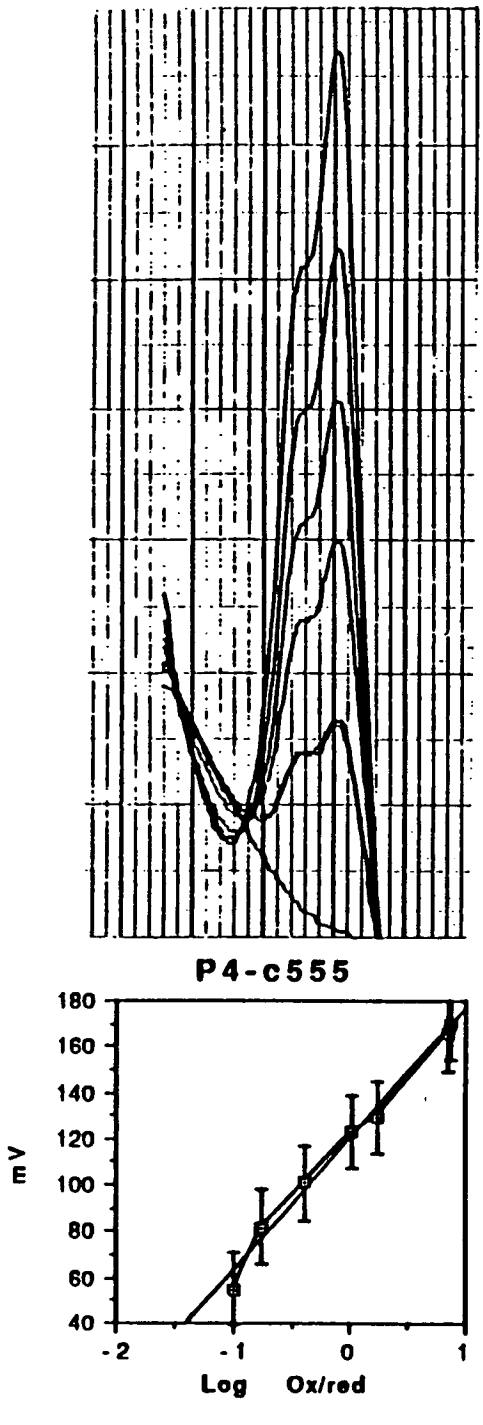


Figure 4.13c. Cytochrome P4-c555. Only the oxidative titration spectra are shown. Slope 57.

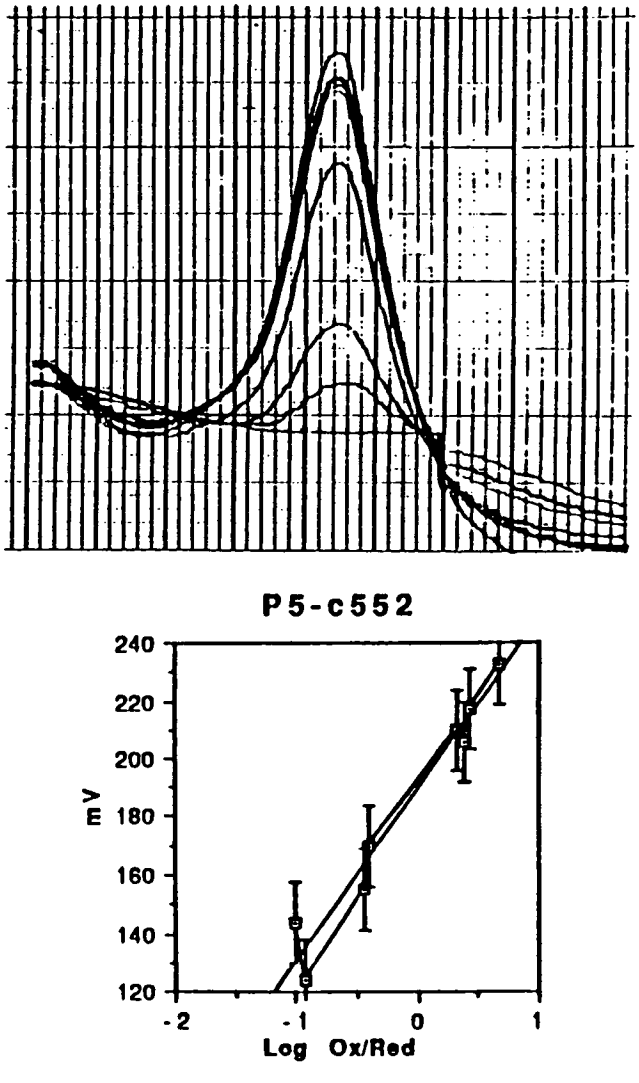


Figure 4.13d. Cytochrome P5-c552. Only the oxidative titration spectra are shown. Slope 59.

Table 4.4. **Absorption maxima, extinction ratios and midpoint redox potentials of cytochromes c from *Bacillus azotoformans*.**

	Reduced spectra, absorption maxima. α nm	β nm	γ nm	280/ $alpha_{red}$	$Soret_{red}$/ $alpha_{red}$	$alpha_{red}$/ $beta_{red}$	Midpoint redox potential mV
P1-c552	552	523	415	0.70	5.61	1.61	122
P2-c551	551	522	416	0.84	6.21	1.69	93
P4-c555	555	523	417	0.72	6.56	1.38	124
P5-c552	552	524	417	0.78	5.14	1.66	188

curves is acceptable for all the isolated cytochrome c polypeptides, except forP2-c551, which is too steep.

4.4. AN IDENTIFICATION OF A B-TYPE CYTOCHROME IN THE MEMBRANE FRACTION OF *B. AZOTOFORMANS*

A red precipitate was observed, if a 1% Triton x100 extract of proteolysed membranes was centrifuged at 31000 g in SS-34 rotor, for approximately 1 hour. A sample of this precipitate was analysed spectrophotometrically and on a 15% SDS polyacrylamid gel. According to SDS-PAGE analysis, the precipitate appeared to consist of a single cytochrome (figure 4.14). This cytochrome may correspond to band 6 in Figure 4.1(apparent mw 22 kD). The absorption spectra of the sample was more complicated (figure 4.15). The alpha maximum indicates, that this is a b-type cytochrome, although it was higher than usually encountered (around 570 nm). Furthermore, this maximum disappeared with time, and another maximum appeared around 560 nm. This happened during a 15 min period, after reduction with dithionite. The spectrum did not have a pronounced β-peak. This spectrum indicates, that there may be more than one b-type cytochrome in the precipitate, with at least one of them retaining its haem during the SDS-PAGE analysis. The precipitate might also contain the cytochrome which appears to be induced during denitrification, but only appeared on a 20% SDS gel (figure 4.16). As protein staining of the electrophoresed sample revealed a complicated protein composition, further purification was not attempted.

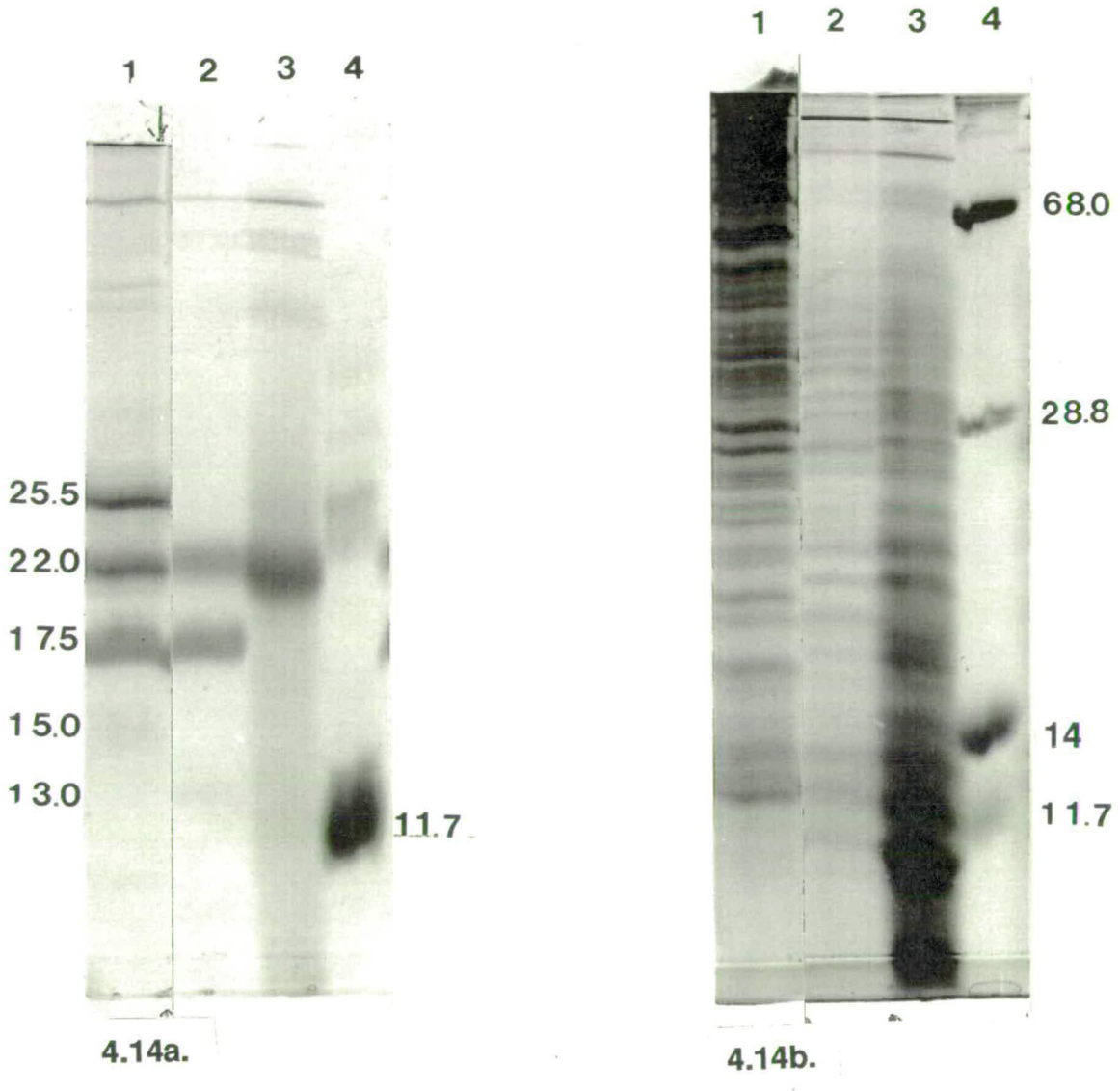
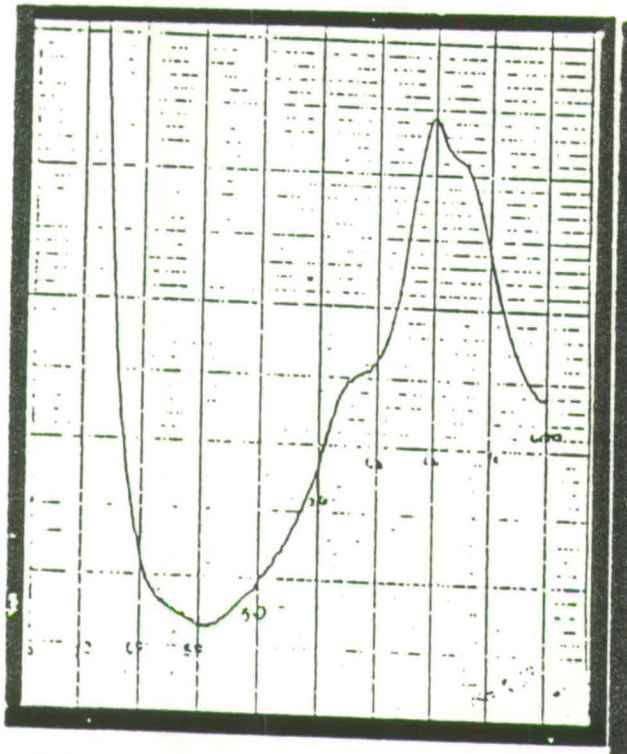
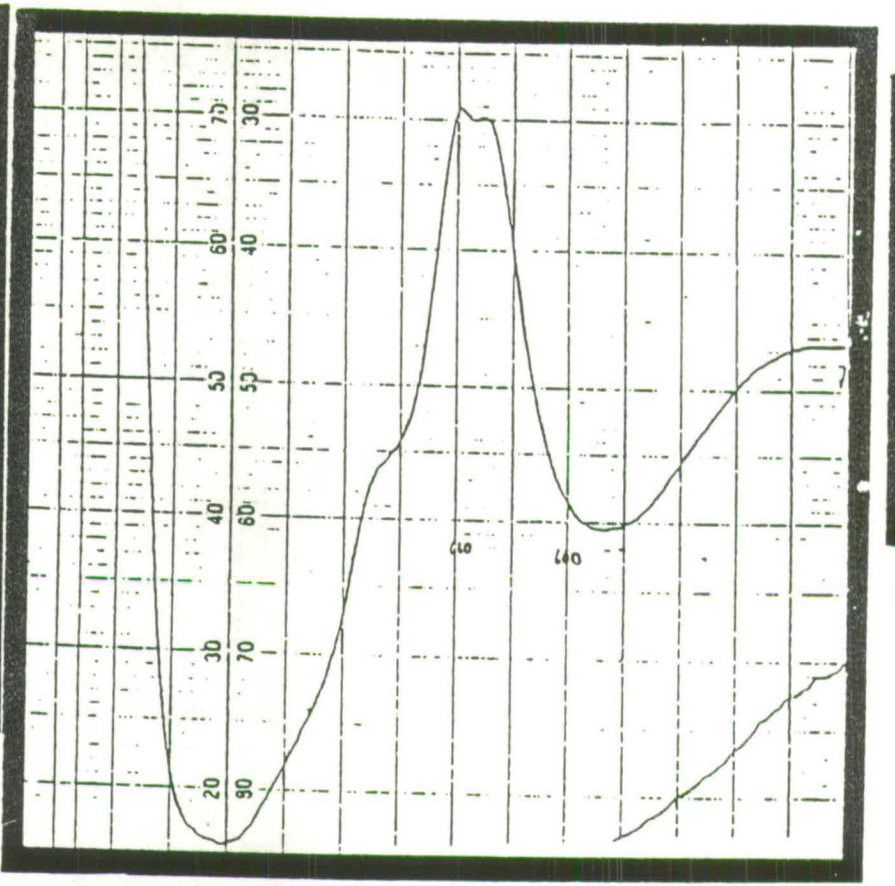


Figure 4.14. 15% SDS-PAGE analyses of the membrane fraction of *B. azotoformans* after trypsin proteolysis and Triton-100 extraction.

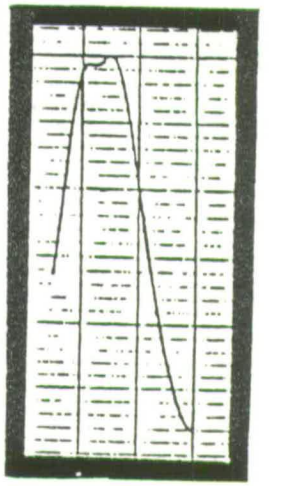
A red precipitate was observed, after prolonged centrifugation of trypsin treated and 2% Triton x100 extracted membranes from *B. azotoformans*, grown under denitrifying conditions. This pellet was resuspended in 0.05 M ammonium acetate, 2% Triton x100, repelled by ultracentrifugation at 100000 g and suspended again in the same buffer. A sample of the solution was analysed by stacking 15% SDS PAGE, and also spectrophotometrically (figure 4.16). *Lane 1*: untreated membranes. *Lane 2*: Washed trypsin treated membranes. *Lane 3*: Trypsin treated and Triton extracted membranes.
Figure 4.14a. Haem staining.
Figure 4.14b. Protein staining.



4.15a.



4.15b.



4.15c.

Figure 4.15. Difference spectra over the alpha and beta peaks of the cytochrome in the membrane fraction of *B. azotoformans*, after trypsin proteolysis and Triton x100 extraction.
The alpha absorption maxima are at approximately at 560 nm and 570 nm.
Figure 4.15a. 1 minute after reduction with sodium dithionite. Only the alpha peak is shown.
Figure 4.15b. 10 minutes after reduction with sodium dithionite. Alpha and beta peaks are shown.
Figure 4.15c. 15 minutes after reduction with sodium dithionite. Only the alpha peak is shown.

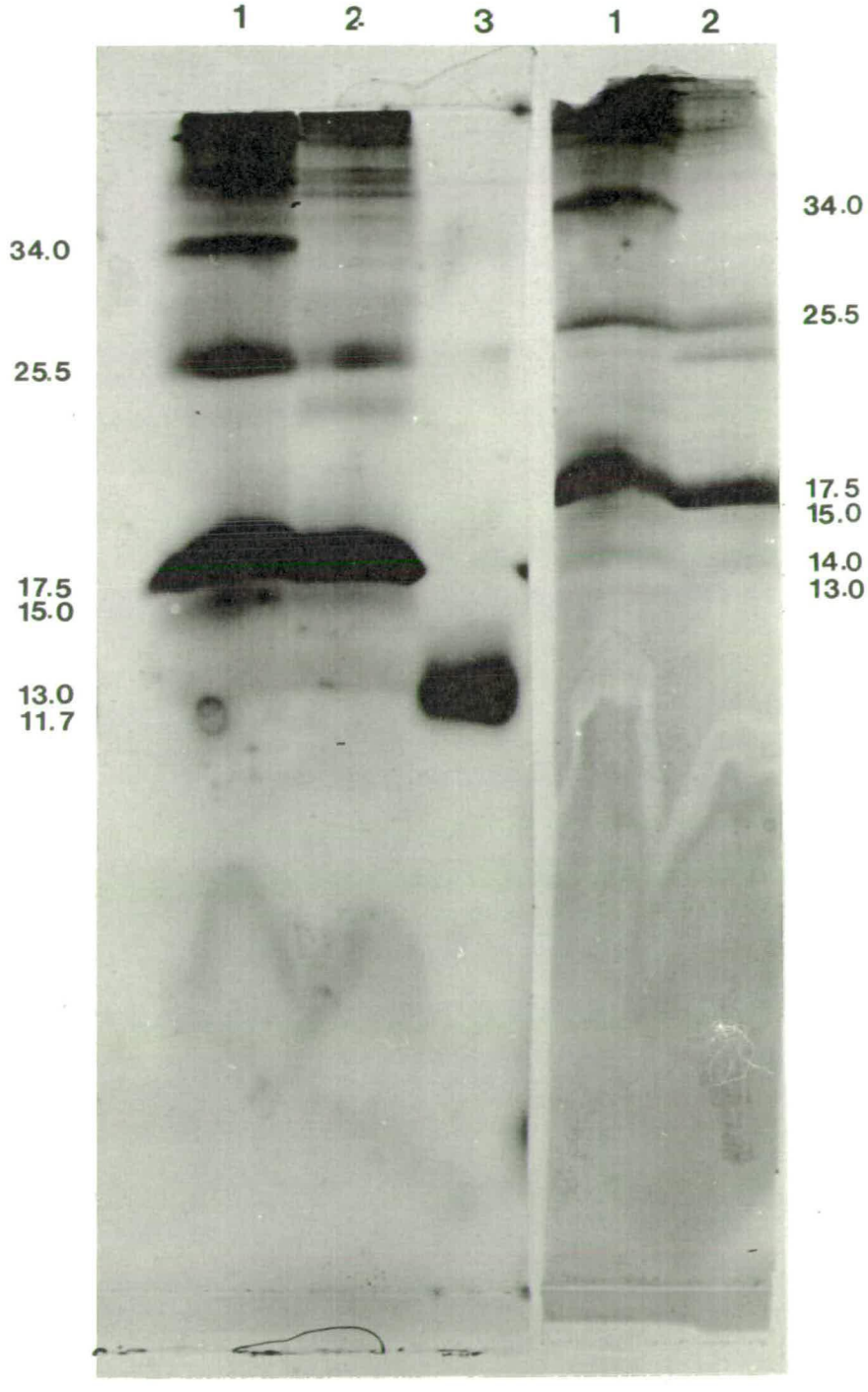


Figure 4.16. 20% SDS-PAGE analysis of the untreated membrane fraction from *B. azotoformans*.
The two haem stained gels represent two separate experiments.
Lane 1: Aerobically grown cells.
Lane 2: Denitrifying culture.
Lane 3: Horse cytochrome c.

4.5 SUMMARY

1) Trypsin, which was used for large scale isolation of cytochrome c polypeptides, released about half of the total cytochrome c content of *Bacillus azotoformans*, measured in Triton membrane extracts of small scale cultures.

2) The yield of trypsin releasable cytochromes was approximately 25%.

3) Four different cytochromes c were isolated; **P1-c552**, **P2-c551**, **P3&P4-c555** and **P5-c552.**

4) As seen by the SDS-PAGE analyses, the purified FPLC fractions consist of different proteolytic fragments. That these fragments derive from the same protein or, that the FPLC fractions consist mainly of one major cytochrome c polypetide, is supported by N-terminal analysis, absorption spectra, redox potential measurements, and amino acid composition analyses. Single N-terminal amino acid may indicate, that the raggedness observed, is at the c terminal end.

6) The main properties of the cytochromes obtrained from *Bacillus azotoformans* are summarised in Table 4.5.

Table 4.5. Data on cytochromes c obtained from *B. azotoformans.*

	Cytochromes				
	P1	P2	P3	P4	P5
Mol. weight (kD) (a. acid composition)	8,9	9,3	11,3	10,3	13,7
(SDS-PAGE)	8,9	<8	<8	<<8	>9 <12
N-terminal amino acid	Gly	Gln/Glu	Gln/Glu	Gln/Glu	Gln/Glu
Absorption maxima (nm, reduced): α	552	551	555	555	552
β	523	522	523	523	524
(Soret) γ	411	409	417	417	412
Extinction ratios 280 nm/α_{red}	0.70	0.84	N.D	0.72	0.78
α_{red}/β_{red}	1.61	1.69	N.D	1.38	1.66
$Soret_{red}/\alpha_{red}$	5.61	6.21	N.D	6.56	5.14
Redox potential	122	93	N.D	124	188
Conditions of expression	Aer*/An*	Aer/A	Aer/An	Aer/An	Aer/An

* Aerobic/Anaerobic growth conditions. *

4.6. CONTINUATION

The following immediate and long-term goals would be interesting to pursue.

1) To aid physiological studies involving cytochromes and the different metabolic activities, the immediate problem is to identify the cytochromes on SDS-PAGE. This may be achieved by raising antibodies to the proteolytic fragments which were isolated, and probing western blots of a SDS-PAGE gel of untreated membrane fractions.

2) To obtain the complete sequences for the cytochromes c which were isolated in this study. The yields of two of the four cytochromes obtained, were good enough to allow manual sequencing, and enough for all of them to be microsequenced. Some sequence information was obtained for all, which may be enough for designing and making oligonucleotide probes.

3) For many reasons (see chapter 6) it would be interesting to classify all cytochrome c ´species´ in a Gram positive organism. According to SDS-PAGE analyses they may be numerous in *B. azotoformans*, and possibly more than the four obtained. Thus different methods for isolation of the remaining ones may be needed. These could involve:

a) different proteases, with different specificity from trypsin, and proceed similarly as was done in this study.

b) two of the cytochromes might be released with relatively mild treatment, such as high salt (2.5 M NaCl) or butanol extraction (these two cytochromes seemed to be hydrophilic enough to show up on a native gel).

c) cytochromes c might be solubilised with detergents before chromatographic separation. Triton x100 solubilised the cytochromes c, which were left after the trypsin treatment. These were mainly the bands 8 and 11. The Triton extract was applied to DEAE column, pH 8.5. The cytochromes stuck to the top of the column and were eluted with salt gradient from 0.0-0.5 M NaCl. They eluted over a great salt concentration range, but still the elutate resolved into two large peaks. Upon examination, it was found, that the composition was more or less the same, in the two peaks, with regard to cytochrome c. However, greater resolution could perhaps be obtained by anionic FPLC exchange, chromatofocusing, or affinity columns.

d) an interesting possibility is preparative electrophoresis. Sufficient amounts could be obtained for direct sequencing or for raising antibodies, and the antibodies could be used to screen genomic libraries for *B. azotoformans* for specific cytochrome clones. Thus, time-consuming purification procedures could be side-stepped.

CHAPTER 5
5. SEQUENCING OF CYTOCHROMES C FROM *BACILLUS AZOTOFORMANS*

Peptides were generated with different proteases. The peptides are designated with a letter which refers to the protease used and a number for distinguishing between the peptides: T=trypsin; Th=thermolysin; V=staphylococcal protease, V8.

Figure 5.1 shows a schematic representation of the procedure for peptide generation and purification for sequencing

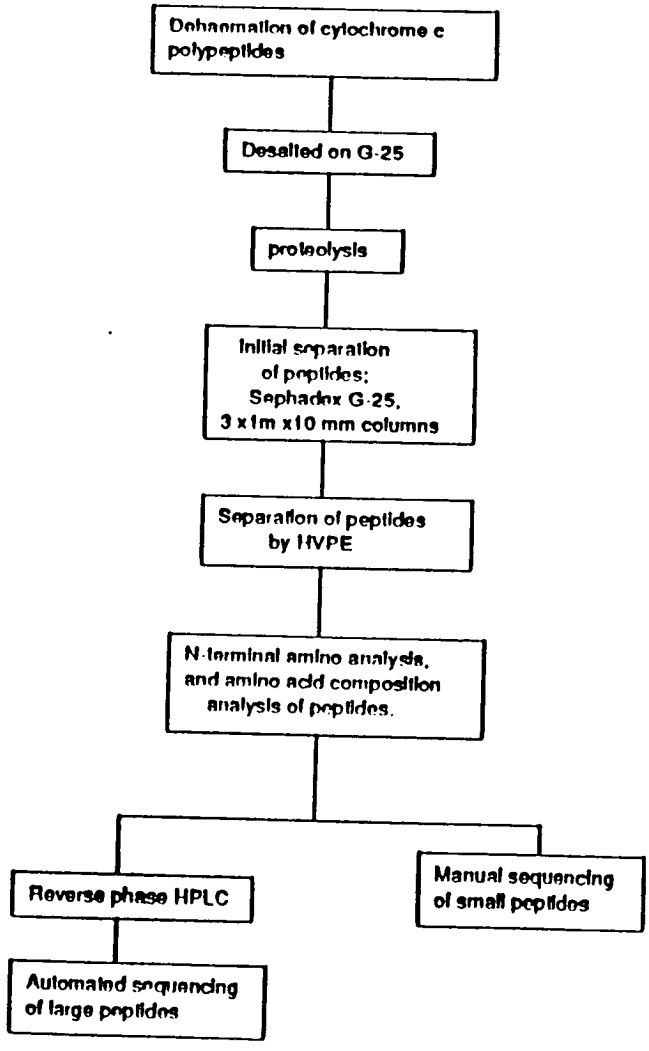


Figure 5.1. A schematic representation of the procedure for the generation and purification of peptides from cytochromes c for sequencing.

5.1. CYTOCHROME P1-C552

Figure 5.2 shows the proposal for the primary structure of the cytochrome P1-c552, the positions of the peptides generated by proteolysis, and the schematic representation of sequencing results and other analyses.

5.1.1. Sequencing of P1-C552 from The N-Terminal End

The FPLC purified protein protein was dehaemed and the apoprotein purified by reverse phase HPLC. A single symmetrical peak was obtained and collected for automated sequencing. The protein sequenced cleanly for the first 21 amino acids, after that the sequence became unreadable. Table A1 in appendix shows the yields of amino acids in each cycle of the sequencing run. Proposed sequence: **1-Gly-Gln-Ala-Thr-Glu-Gln-Ala-Ala-Pro-Ala-Ala-Asp-Asp-Pro-Glu-Gly-Met-Val-Lys-Ala-Ser-21.**

5.1.2. Peptides, Generated by Trypsin Proteolysis

Figure 5.3 shows the electrophoresis map of distribution of peptides in the G25 gel filtration eluate and the corresponding 280 nm absorption trace. Tables 5.1 and 5.2 show, respectively, the amino acid composition, and a summary of other relevant information.

Peptides generated with trypsin and isolated, covered the entire sequence. All lysyl and arginyl bonds were hydrolysed, although the lys-69 bond was only partially cleaved. The occurrence of one peptide (T6) can only be explained by partial cleavage of the trp-76 bond. This may be the result of an acid hydrolysis of the bond in the presence of formic acid (pH 5.1) during the G25 gel filtration, but more likely it is because of chymotrypsin activity in the trypsin. Peptides were oxidised after the gel filtration. The HVPE pH 6.5 electrophoresis map (figure 5.3) of the gel filtration eluate was visualised under UV before ninhydrin/collidine staining, but no fluorescence was detected. The yields were not very high, on the average approximately 10% of the starting amount (1400 nmol).

Comments on individual peptides are below. Residue designations, starting with a capital letter, signify residue identities obtained by sequencing of that particular peptide. Residue designations starting with a

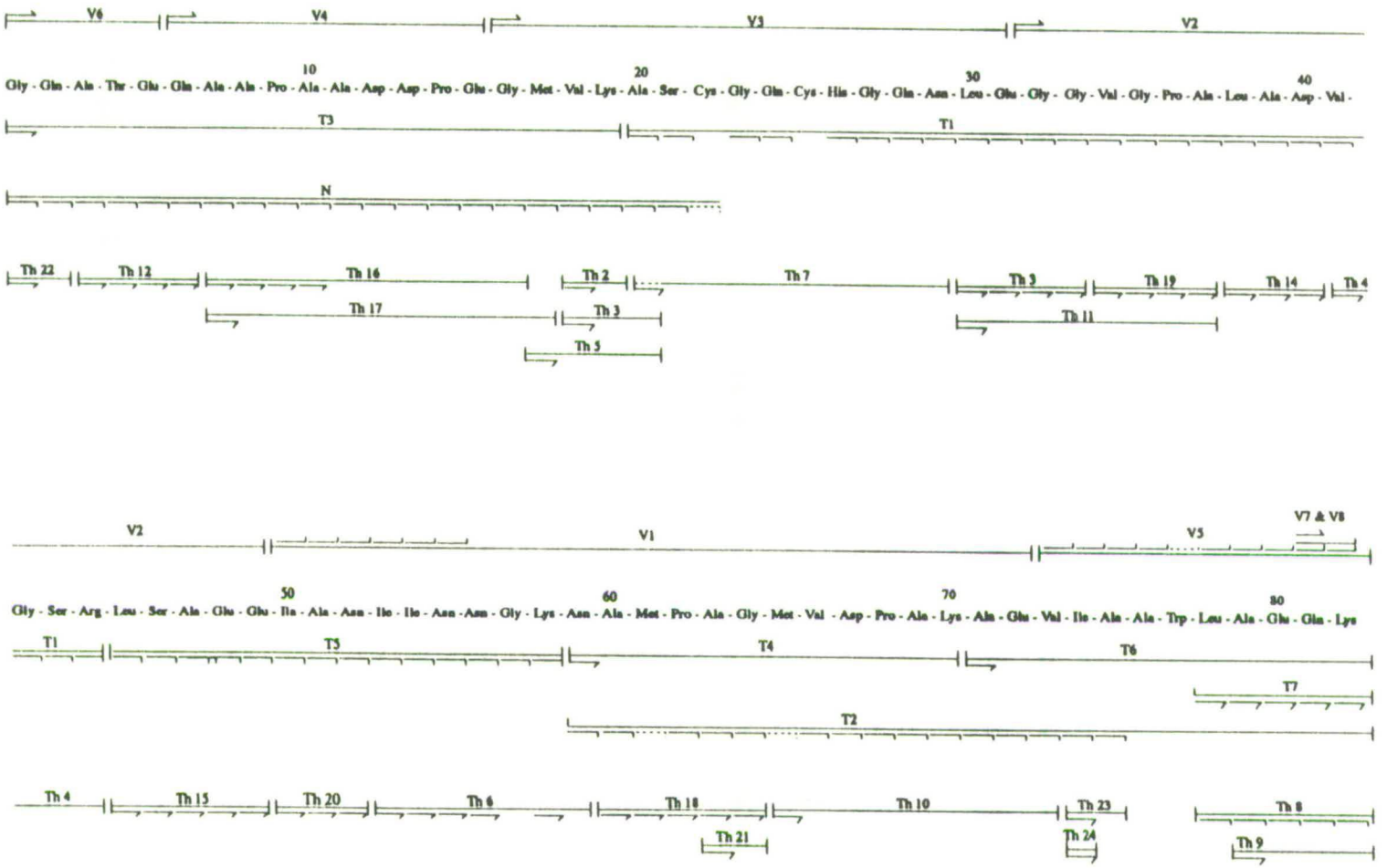


Figure 5.2. Amino acid sequence of *Bacillus azotoformans* cytochrome P1-c552.

The peptides are shown in the proposed sequence order for the cytochrome P1-c552 N: N-terminus sequencing. T: Peptides derived by digestion with trypsin. Th: Peptides derived by digestion with thermolysin. V: Peptides derived by digestion with staphylococcal protease, V8. Continuous lines indicate quantitative amino acid analyses. Under peptide lines, ⟶ indicates end groups and subsequent residues revealed by phenyl isothiocyanated degradation and identified by the dansyl method, ⟶ indicates residues obtained by automated sequencing. Broken lines indicate substandard results or a residue unrecognised by the sequenators program.

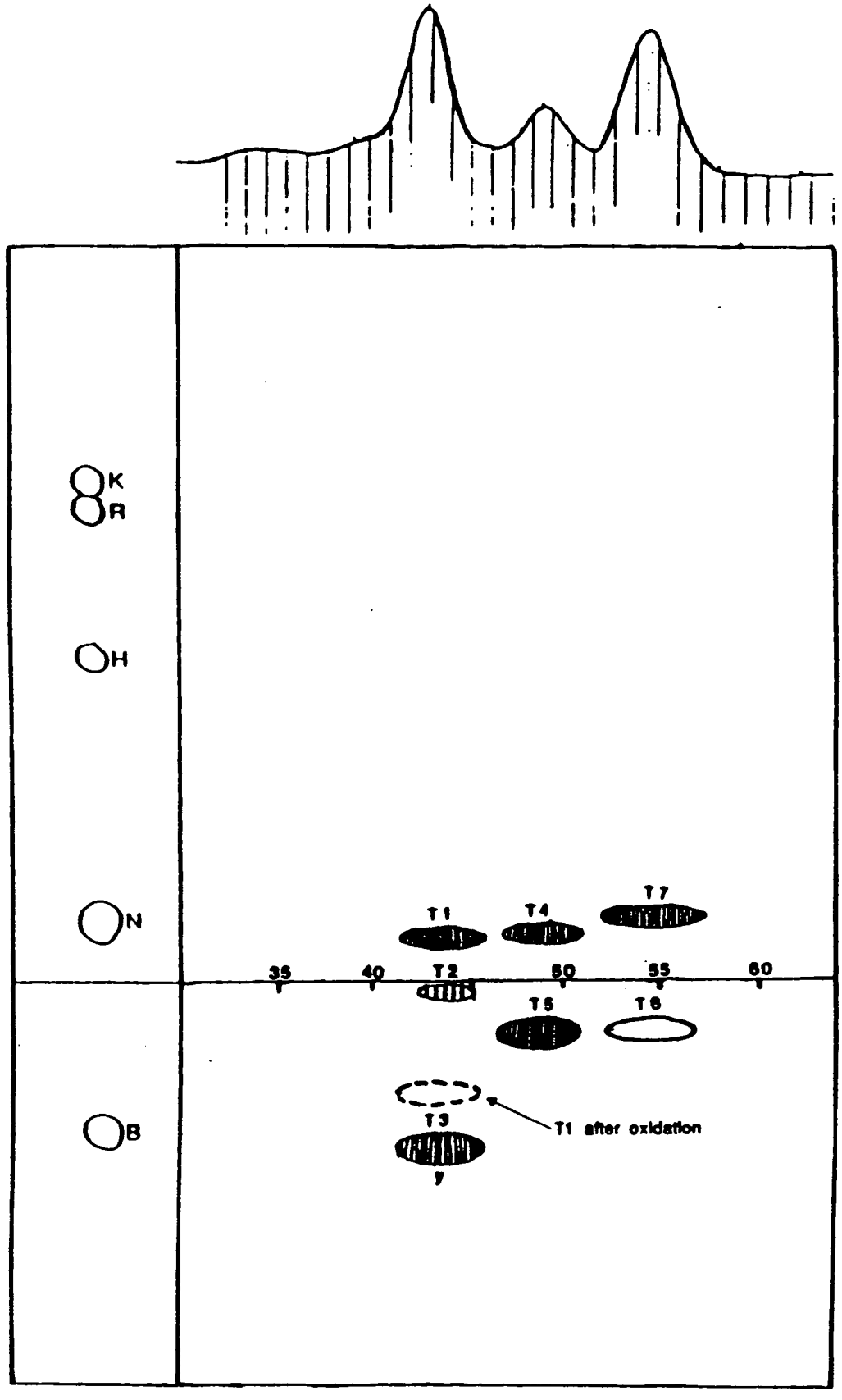


Figure 5.3. HVPE, pH 6.5 mapping of tryptic peptides fractionated by G25 gel filtration. The corresponding 280 nm absorption trace of the gel elution is shown on top.

Table 5.1. Amino acid composition of tryptic peptides from the P1-c552 apocytochrome.

Amino acid	T1	T2	T3	T4	T5	T6	T7
Lys	(0)	1.70 (2)	1.06 (1)	0.72 (1)	1.05 (1)	0.53 (1)	1.00 (1)
His	0.86 (1)	(0)	(0)	(0)	(0)	(0)	(0)
Arg	0.67 (1)	(0)	(0)	(0)	(0)	(0)	(0)
Asx	1.95 (2)	2.28 (2)N	2.14 (2)	1.60 (2)N	3.00 (3)	(0)	(0)
Thr	(0)	(0)	0.76 (1)	(0)	(0)	(0)	(0)
Ser	2.81 (3)	1.50* (0)	0.97* (0)	0.74* (0)	0.88 (1)	4.03 (4)n	(0)
Glx	2.05 (2)	3.06 (3)	4.23 (4)	(0)	1.97 (2)	(0)	2.09 (2)
Pro	0.86 (1)	1.71 (2)	1.67 (2)	1.42 (2)	(0)	(0)	(0)
Gly	5.14 (5)	2.26* (1)	2.37 (2)N	2.46* (1)	1.18 (1)	0.88* (0)	(0)
Ala	3.93 (4)N	6.60 (7)	5.25 (5)	2.86 (3)	2.34 (2)	3.43 (3)	1.01 (1)
Val	1.85 (2)	2.27 (2)	0.86 (1)	1.00 (1)	(0)	1.61 (1)	(0)
Met	(0)	1.70 (2)	0.45 (1)	1.53 (2)	(0)	(0)	(0)
Ile	(0)	1.38 (1)	(0)	(0)	1.90* (3)	1.57 (1)	(0)
Leu	2.00 (2)	1.61 (2)	(0)	(0)	0.89 (1)N	1.51 (1)	1.03 (1)N
Tyr	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Phe	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Cys	1.82 (2)	(0)	(0)	(0)	(0)	(0)	(0)
Trp	(0)	(1)	(0)	(0)	(0)	*(1)	(0)
Total	25	24	19	12	14	12	5
Seq. pos.	A20-R44	N59-K82	G1-K19	N59-K70	L45-K58	A71-K82	L79-K82

Amounts used for Beckman 120C analysis (peptides T1, T2, T3, T4, T5 and T7): 5-20 nmols.
Amount used for amino acid analysis on the Applied Biosystem model 420A (peptide T6) 20 pmol.
Values less than 0.1 are omitted. Values in parenthesis are deduced from the sequence.
Values marked with an asterix (*) are incongruent with deduced sequence.
N: designates a N-termnial amino acid.
+/++/+++ : qualitative amino acid analysis only.
Comments on individual peptides are in text.

Table 5.2. General information on tryptic peptides from P1-c552.

Peptide	Purification[a]	V/Vo (G25)	Mobility[b]	Yield[c] nmol %	N-term. a.acid	number a.acids	position
T1	G25,6.5 HPLC	1.09	-0.01	365 24	Ala	25	20-44
T2	G25,6.5	1.07	-0.23	142 10	Asx	24	69-82
T3	G25,6.5	1.10	-0.58	200 13	Thr	19	1-19
T4	G25,6.5 HPLC	1.23	0.00	114 8	Asx	12	59-70
T5	G25,6.5 HPLC	1.23	-0.33	174 12	Leu	14	45-58
T6	G25,6.5	1.38	-0.32	low	Ala	12	71-82
T7	G25,6.5	1.35	0.00	109 7	Leu	5	79-82

[a] G25: gel filtration; High Voltage Paper Electrophoresis at pH 6.5; HPLC: High pressure liquid chromatography.
[b] mobility in pH 6.5 HVPE relative to lysine(+1).
[c] % : percentage recovery of protein digested (1500 nmol).

lower case letter signify residue identities obtained by amino acid composition analysis, only. When residues are connected with a hyphen, it signifies that they are in the order which was determined by sequencing of the corresponding peptide or some other peptide from the same region.

T1: This is the haem peptide as indicated by the change from neutral to acidic mobility on HVPE pH 6.5 after oxidation (figure 5.3), reflecting the change of the cysteines to cysteic acids. The peptide was stained for histidine and its presence was confirmed by the Pauly test (see chapter 2). The peptide was completely sequenced by automated sequencing. Table A2 in appendix A shows the yields of amino acids in each cycle of the sequencing run. Proposed sequence: **20-Ala-Ser-Cys-Ala-Ser-Gly-Gln-Cys-His-Gly-Gln-Asn-Leu-Glu-Gly-Gly-Val-Gly-Pro-Ala-Leu-Ala-Asp-Val-Gly-Ser-Arg-44.**

T2: This peptide was isolated in a comparatively low yield. The amino acid analysis was originally incorrectly interpreted and the peptide was thought to contain only one lysine. This led to an underestimation of its size. The automated sequencing was interrupted prematurely and only the first 17 residues were sequenced. The recalculated amino acid composition of the peptide conforms with it being a C-terminal peptide from asn-59 to lys-82 containing two lysines. It was heavily contaminated with serine and glycine. These amino acids are commonly observed as contaminants in peptide preparations. The methionines did not show up, but a prominent peak was observed in the chromograms for the corresponding cycles. It did not relate to any of the amino acid peaks in chromogram for the amino acid standards. According to amino acid analysis and manual sequencing of smaller thermolytic peptides from the same region, this peak is probably methionine sulphone. The elution time of this peak was 9.18 min and is marked into the elution trace of cycle 3 in figure A1, appendix A. Table A3 in appendix A shows the yields of amino acids in each cycle of the sequencing run. The comparatively high absorption peak (280 nm) for the corresponding fraction of the gel filtration eluate is probably due to the tryptophan in this peptide. Proposed sequence: **59-Asn-Ala-(met)-Pro-Ala-Gly-(met)-Val-Asp-Pro-Ala-Lys-Ala-Glu-Val-Ile-Ala-(ala-trp-leu-ala-glu-gln-lys)-82.**

T3: The yellow spot after preparative HVPE pH 6.5 and ninhydrin/collidine staining, suggests that glycine was the N-terminal amino acid. This was

subsequently confirmed by N-terminal amino acid analysis. The peptide was not sequenced as its amino acid composition accorded with the N-terminal sequence, already obtained from the dehaemed protein. The peptide was contaminated with serine. The great acidic mobility (figure 5.3, table 5.2) indicates that there are quite a few aspartate and/or glutamate residues in the peptide. The proposed sequence: **1-gly-glx-ala-thr-glx-glx-ala-ala-pro-ala-ala-asx-asx-glx-gly-met-val-lys-19.**

T4: This peptide showed up as a grey spot by ninhydrin/collidine staining on the HVPE paper. The peptide was not sequenced as the sequence had already been obtained from the T2 peptide. The preparation was contaminated with glycine. The neutral mobility (figure 5.3, table 5.2) suggest one aspartate and one aspargine. The proposed sequence: **59-asx-ala-met-pro-ala-gly-met-val-asx-pro-ala-lys-70.**

T5: This peptide was sequenced by automated sequencing. The point worth noting is, that the amino acid composition shows only two isoleucines, where in reality it has three. This discrepancy was observed for all peptides containing this sequence part subsequently obtained by treatment with other proteases. This may be explained by the fact that isoleucines form peptide bonds that are particularly resistant to hydrolysis. Table A4 in appendix A shows the yields of amino acids in each cycle of the sequencing run. Proposed sequence: **45-Leu-Ser-Ala-Glu-Glu-Ile-Ala-Asn-Ile-Ile-Asn-Asn-Gly-Lys-58.**

T6: This peptide was obtained in a very low yield, which is not surprising as the same sequence is found (wholly and partially) in two other peptides. The amino acid composition analysis was performed on Applied Biosystems, model 420A. Results obtained with this apparatus were notably more inaccurate than the ones usually obtained on the Beckman 120C analyser. The acidic mobility (net charge -1), suggests two glutamates. Sequencing was not attempted, but the proposed sequence is: **71-ala-glx-val-ile-ala-ala-trp-leu-ala-glx-glx-lys-82.**

T7: This peptide may have been generated by acid hydrolysis of the trp-76 bond during the gel filtration, but more likely, it was formed by chymotrypsin activity in the trypsin. The high absorption peak (280 nm) for the fractions containing this peptide is probably due to the presence

of tryptophan in both the T6 and T7 peptides. The neutral mobility (0.00) suggests one glutamate and one glutamine. The yield of the T7 peptide was sufficient to allow manual sequencing. The proposed sequence: **78-Leu-Ala-Glx-Glx-Lys-82.**

5.1.3. Peptides, Generated by Thermolysin Proteolysis

Figure 5.4 shows the electrophoresis map of the distribution of peptides in the G25 gel filtration eluate and the corresponding 280 nm absorption trace. Tables 5.3 and 5.4 show, respectively, the amino acid composition of the individual peptides, and a summary of other relevant information.

Peptides were isolated that covered the greater part of the protein molecule.

Comments on individual peptides are below. Residue designations starting with a capital letter signify residue identity obtained by sequencing of that particular peptide. Residue designations starting with a lower case letter signify residue identities obtained by amino acid composition analysis only. When residues are connected with a hyphen, it signifies that they are in the order which was determined by sequencing of the corresponding peptide, or some other peptide from the same region.

Th1: The spot was very weak and an elution was not attempted.

Th2: Only qualitative amino acid analysis and N-terminal analysis were performed. The proposed sequence: **18-Val-(lys)-19.**

Th3: The peptide was sequenced manually. The proposed sequence: **18-Val-Lys-Ala-20.**

Th4: Only qualitative amino acid and N-terminal analyses were performed.

The proposed sequence: **41-Val-(gly-ser-arg)-44.**

Th5: Amino acid analysis was done on Applied Biosystems, model 420A. Methionine was identified as the N-terminal amino acid, but it did not show up in the amino acid analysis. It may have been oxidised to methionine sulphone, as it did not elute at the expected time. In support of this, was the presence of a peak with the retention time 7.5 min in the

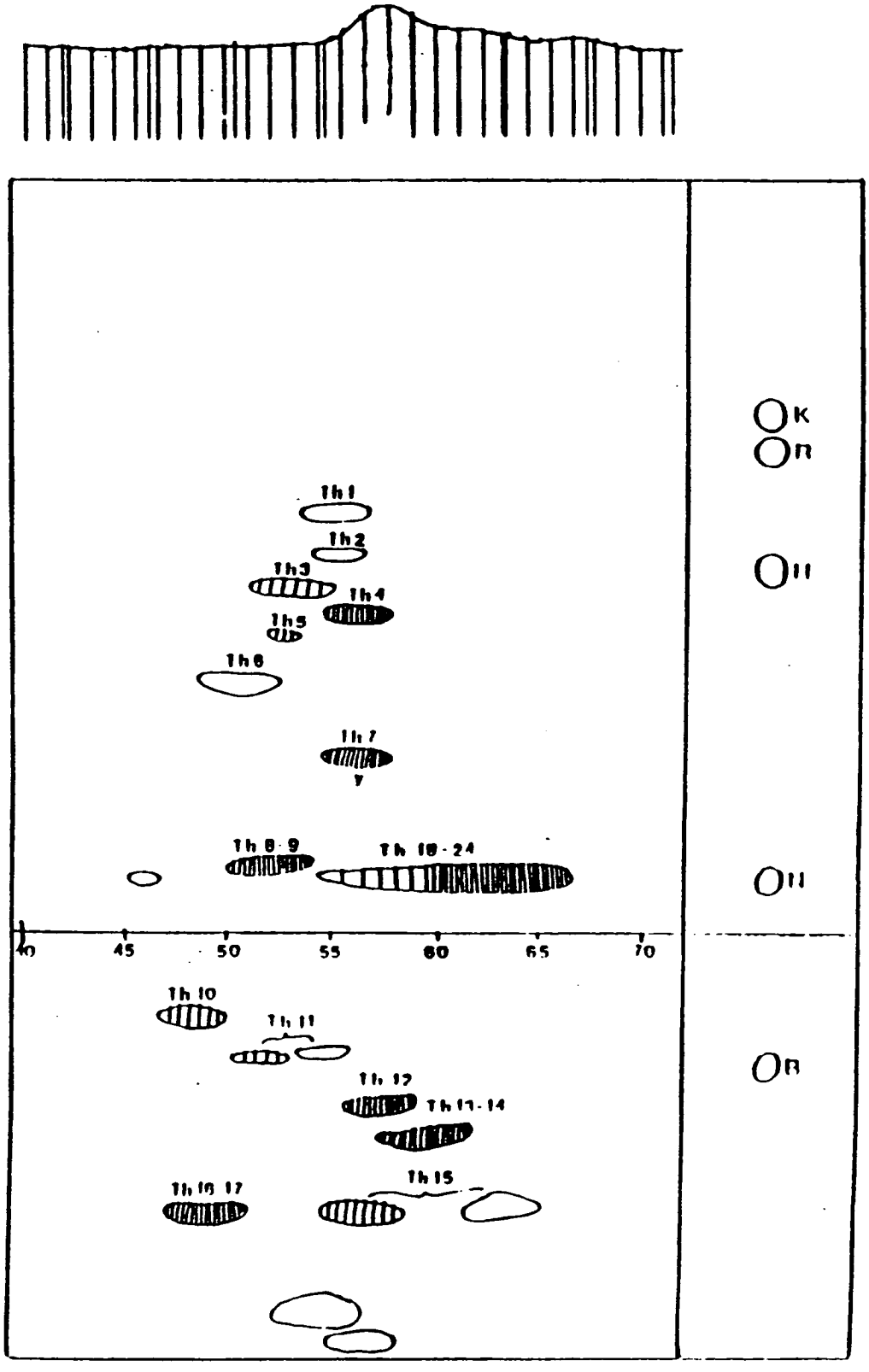


Figure 5.4. HVPE, pH 6.5 mapping of thermolytic peptides, fractionated by G25 gel filtrations. The corresponding 280 nm absorption trace of the gel filtration elution is shown on top.

Table 5.3. Amino acid composition of peptides (Th1-Th24) from P1-c552 apocytochrome generated with thermolysin.

Amino acid	Th2	Th3	Th4	Th5*	Th6	Th7	Th8	Th9
Lys	++ (1)	1.00 (1)	(0)	1.00 (1)	0.90 (1)	(0)	0.96 (1)	++ (1)
His	(0)	(0)	(0)	(0)	(0)	++ (1)	(0)	(0)
Arg	(0)	(0)	++ (1)	(0)	(0)	(0)	(0)	(0)
Asx	(0)	(0)	(0)	(0)	3.01 (3)	++ (1)	(0)	(0)
Thr	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Ser	(0)	(0)	++ (1)	(0)	(0)	++ (1)N	0.75* (0)	(0)
Glx	(0)	(0)	(0)	0.21* (0)	(0)	++ (1)	2.00 (2)	++ (2)
Pro	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Gly	(0)	(0)	++ (1)	(0)	1.09 (1)	++ (1)	0.46* (0)	(0)
Ala	(0)	1.12 (1)	(0)	1.09 (1)	(0)	++ (1)	1.09 (1)	++ (1)N
Val	++ (1)N	1.30 (1)N	++ (1)N	1.11 (1)	(0)	(0)	(0)	(0)
Met	(0)	(0)	(0)	*(1)N	(0)	(0)	(0)	(0)
Ile	(0)	(0)	(0)	(0)	1.01* (2)N	(0)	(0)	(0)
Leu	(0)	(0)	(0)	(0)	(0)	(0)	0.99 (1)N	(0)
Tyr	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Phe	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Cys	(0)	(0)	(0)	(0)	(0)	++ (1)	(0)	(0)
Trp	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Total	2	3	4	4	7	7	5	4
Seq. pos.	V18-K19	V18-A20	V41-R44	M27-A20	A53-N59	S21-N29	L78-K82	L79-K82

Table 5.3. cont. Amino acid composition of peptides from the P1-c552 apocytochrome generated with thermolysin.

Amino acid	Th10*	Th11	Th12	Th13	Th14	Th15	Th16	Th17
Lys	0.97 (1)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
His	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Arg	0.29 (0)	++ (0)	(0)	(0)	(0)	(0)	(0)	(0)
Asx	0.59 (1)	(0)	(0)	(0)	1.02 (1)	(0)	1.83 (2)	1.37* (2)
Thr	0.11 (0)	(0)	1.00 (1)	(0)	(0)	(0)	(0)	(0)
Ser	(0)	(0)	(0)	(0)	(0)	0.96 (1)	(0)	(0)
Glx	1.19* (1)	+++ * (1)	2.04 (2)	1.06 (1)	(0)	2.20 (2)	0.99 (1)	1.08 (1)
Pro	1.02 (1)	++ (1)	(0)	(0)	(0)	(0)	1.77 (2)	2.47* (2)
Gly	0.18 (0)	+++ (3)	(0)	2.09 (2)	(0)	(0)	1.01 (1)	1.20 (1)
Ala	2.00 (2)	+++ (1)	1.21 (1)	(0)	0.98 (1)	1.10 (1)	3.92 (4)N	4.00 (4)N
Val	1.05* (2)	++ (1)	(0)	(0)	(0)	(0)	(0)	(0)
Met	0.56 (1)N	(0)	(0)	(0)	(0)	(0)	(0)	0.63 (1)
Ile	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Leu	0.19 (0)	++ (1)N	(0)	0.94 (1)N	1.00 (1)N	0.94 (0)N	(0)	(0)
Tyr	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Phe	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Cys	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Trp	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Total	8	8	4	4	3	5	10	11
Seq. pos.	M65-V73	L30-A37	A3-Q6	L30-G33	L38-D40	L45-E49	A7-G16	A7-M17

Table 5.3. cont. Amino acid composition of peptides from the P1-c552 apocytochrome generated with thermolysin.

Amino acid	Th18	Th19	Th20	Th21	Th22	Th23	Th24
Lys	(0)	(0)	(0)	(0)	(0)	(0)	(0)
His	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Arg	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Asx	(0)	(0)	0.90 (1)	(0)	(0)	(0)	(0)
Thr	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Ser	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Glx	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Pro	2.01 (1)	0.92 (1)	(0)	(0)	1.00 (1)	(0)	(0)
Gly	1.22 (1)	1.04 (1)	(0)	0.87 (1)	1.00 (1)N	(0)	(0)
Ala	0.99 (2)N	1.05 (1)	1.08 (1)	1.13 (1)N	(0)	0.98 (1)	(0)
Val	(0)	1.00 (1)N	(0)	(0)	(0)	(0)	(0)
Met	0.80 (1)	(0)	(0)	(0)	(0)	(0)	(0)
Ile	(0)	(0)	1.10 (1)N	(0)	(0)	1.02 (1)N	+++ (2)N
Leu	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Tyr	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Phe	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Cys	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Trp	(0)	(0)	(0)	(0)	(0)	(0)	? (1)
Total Seq.	5	4	3	2	2	2	3
pos.	A60-G64	V34-A37	I50-N52	A63-G64	G1-Q2	I74-A75	A75-W77?

Amounts used for Beckman 120C analysis: 5-20 nmol. Amount used for amino acid analysis on the Applied Biosystem model 420A: 20 pmol.
Values less than 0.1 are omitted. Values in parenthesis are deduced from the sequence. Comments on individual peptides are in text. Values marked with an asterix (*) are incongruent with deduced sequence.
*: analysis done on the Applied Biosystem model 420A. N: designates a N-terminal amino acid.
+/++/+++ : qualitative amino acid analysis only.

Table 5.4. General information on thermolytic peptides from cytochrome P1-c552.

Peptide	Purification[a]	V/Vo (G25)	Mobility[b]	Yield[c] nmol %	N-term. a.acid	number a.acids	position
Th2	G25:6.5:3.5	1.35	0.69	low	Val	2	18-19
Th3	G25:6.5:3.5	1.31	0.62	113 8	Val	3	18-20
Th4	G25:6.5:3.5	1.38	0.57	>100 >7	Val	4	41-44
Th5	G25:6.5:3.5	1.29	0.52	low	Met	4	17-20
Th6	G25:6.5:3.5	1.25	0.43	260 19	Ala	7	53-59
Th7	G25:6.5	1.38	0.27	low	Ser	7	21-29
Th8	G25:6.5:3.5	1.30	0.02	240 17	Leu	5	78-82
Th9	G25:6.5:3:5	1.30	0.02	low	Leu	4	79-82
Th10	G25:6.5	1.19	-0.30	low	Met	8	65-73
Th11	G25:6.5	1.25/1.33	-0.36	low	Leu	8	30-37
Th12	G25:6.5:3.5	1.40	-0.47	291 21	Ala	4	3-6
Th13	G25:6.5:3.5	1.46	-0.55	259 19	Leu	4	30-33
Th14	G25:6.5:3:5	1.46	-0.55	117 8	Leu	3	38-40
Th15	G25:6.5:3.5	1.39/1.56	-0.69	190 14	Leu	5	45-49
Th16	G25:6.5:3.5	1.21	-0.69	120 9	Ala	10	7-16
Th17	G25:6.5:X[d]	1.21	-0.69	low	Ala	11	7-17
Th18	G25:6.5:X	1.30-1.58	0.00	201 14	Ala	5	60-64
Th19	G25:6.5:X	1.30-1.58	0.00	160 11	Val	4	34-37
Th20	G25:6.5:X	1.30-1.58	0.00	275 20	Ile	3	50-52
Th21	G25:6.5:X	1.30-1.58	0.00	>100 >7	Ala	2	63-64
Th22	G25:6.5:X	1.30-1.58	0.00	80 6	Gly	2	1-2
Th23	G25:6.5:X	1.30-1.58	0.00	70 5	Ile	2	74-75
Th24	G25:6.5:X	1.30-1.58	0.00	low	Ala	3	75-77

[a] G25: gel filtration; 6.5: HVPE pH 6.5; 3.5: HVPE pH 3.5.
[b] mobility in pH 6.5 HVPE relative to mobility of lysine.
[c] % : percentage recovery of starting amount of protein (1400 nmol).
[d] X : descending paper chromatography.

elution trace from the amino acid analyser. The peptide preparation was contaminated with glycine. The proposed sequence is: 17-Met-(val-lys-ala)-20.

Th6: This was a problematic peptide. It showed only one isoleucine in the amino acid composition analysis, but subsequent sequencing revealed two. (Every peptide containing these two isoleucine residues gave this result). Isoleucine forms bonds, particularly resistant to hydrolysis. The peptide showed up in three different spots in repeated preparative HVPE pH 3.5 runs. Aspargine residues next to glycine are known to undergo deamidation via cyclic imide to a mixture of aspα-gly and aspβ-gly (Ambler, 1963) which run the same on HVPE pH 6.5, but separate into three spots on HVPE pH 3.5 as peptides containing asn-gly, aspα-gly and aspβ-gly.

The first five residues were sequenced manually. The glycine did not show up but may have been covered by the DNS-OH, which is coloured blue and runs to approximately same position. Still, it was not detected when the blue spot was cut out, sewn onto a sheet of paper and subjected to descending paper chromatography. The proposed sequence: **53-Ile-Ile-Asx-Asx-(gly)-Lys-(asx)-59.**

Th7: The yellow colour of the peptide spot, after HVPE pH 6.5 and ninhydrin/collidine staining, suggests serine. This was not confirmed by the N-terminal analysis, which showed alanine. Thermolysin is not reported to cut before serine, but it cannot be ruled out as activity may be affected by neighbouring amino acids, in this case cysteine. The amino acid composition analysis was qualitative, only, due to low yield. The positive mobility (the histidine) suggests that there are no glutamates or aspartates in the peptide. The proposed sequence: **20-ala-(ser-cys-gly-glx-cys-his-gly-glx-asx)-29.**

Th8: This peptide was completely sequenced, manually. It covers the same region as T7 and Th9. The neutral mobility of the peptide (figure 5.4, table 5.4) suggests one glutamine and one glutamate. The proposed sequence is: **78-Leu-Ala-Glx-Glx-Lys-82.**

Th9: This is a low yield peptide which ran with Th8 in HVPE pH 6.5, but separated from the Th8 peptide in the preparative HVPE pH 3.5. Only qualitative amino acid analysis was performed. The peptide appears to be

one amino acid shorter than Th8, which is supported by the N-terminal amino acid analysis, which gave alanine. The neutral (figure 5.4, table 5.4) mobility suggests one glutamate and one glutamine. The proposed sequence: **79-Ala-(glx-glx-lys)-82.**

Th10: The amino acid composition analysis was performed on Applied Biosystmes, model 420A, and the peptide appears to be contaminated (see table 5.3). However, the larger integral values add up and its positioning is fairly secure. Methionine was the N-terminal amino acid and the acidic mobility suggests a net charge of -1 and therefore one aspartate and one glutamate. The proposed sequence: **65-Met-(val-asx-pro-ala-lys-ala-glx-val)-73.**

Th11: This was a mixture of two low yield peptides. Only one of them was recovered in a sufficient yield, in the following separation step by HVPE pH 3.5, to allow analysis. Only qualitative amino acid and N-terminal amino acid analyses were performed. Leucine is the N-terminal amino acid. The strength of the glutamate spot, noted in the qualitative amino acid analysis, is stronger than the deduced sequence gave reason to expect, but the estimations of spot intensities are subjective and conclusions drawn thereof are at best inaccurate. The acidic mobility suggests glx = glutamate. Shorter peptides that also covered this region (**Th13, Th14** and **Th19**) were isolated and sequenced. The proposed sequence is: **30-Leu-(glx-gly-gly-val-pro-ala)-36.**

Th12: This peptide was sequenced manually. The acidic mobility suggests one glutamate and one glutamine. The greater mobility of this peptide compared with Th11 is due to the smaller size of the former. The proposed sequence: **3-Ala-Thr-Glx-Glx-6.**

Th13: Peptides **Th13** and **Th14** ran together in HVPE pH 6.5 and were separated by HVPE pH 3.5, mobility suggests glutamate. This peptide was completely sequenced, manually. The proposed sequence is: **30-Leu-glu-Gly-Gly-33.**

Th14: See also **Th13**. This peptide was sequenced manually. The acidic mobility suggests aspartate. The proposed sequence: **38-Leu-Ala-Asx-40.**

Th15: This is a mixture of two peptides. One of them was not recovered in sufficient yield by the preparative HVPE pH 3.5, to allow analysis. The other one was sequenced, manually. The acidic mobility suggests the presence of two glutamates. The proposed sequence: **45-Leu-Ser-Ala-Glx-Glx-49.**

Th16: This peptide was partially sequenced, manually. The protein region in which it is positioned, had been completely and convincingly sequenced from the N-terminal of undigested protein. The proposed sequence: **7-Ala-Ala-Pro-Ala-(ala-asx-asx-pro-glx-gly)-16.**

Th17: This peptide was not sequenced. It copurified with **Th16** and was separated from it by HVPE pH 3.5. According to the amino acid and N-terminal analysis, these peptides are identical, apart from an additional residue (methionine) at the C-terminal in Th17. The proposed sequence: **7-Ala-(ala-pro-ala-ala-asx-asx-pro-glx-gly-met)-17.**

Th18: This peptide was sequenced, manually. The proposed sequence **60-Ala-Met-Pro-Ala-Gly-64.**

Th19: This peptide was sequenced, manually. The proposed sequence: **34-Val-Gly-Pro-Ala-37.**

Th20: This peptide was sequenced, manually. The neutral mobility suggests aspargine. The proposed sequence: **50-Ile-Ala-Asx-52.**

Th21: Quantitative amino acid composition and N-terminal amino acid analyses were performed. The proposed sequence: **63-Ala-(gly)-64.**

Th22: Quantitative amino acid composition and N-terminal amino acid analysis were performed. The proposed sequence: **1-Gln-(gly)-2.**

Th23: Quantitative amino acid composition and N-terminal amino acid analysis were performed. The proposed sequence: **74-Ile-(ala)-75.**

Th24: Only qualitative amino analysis was performed. The proposed sequence: **53-(ile).**

5.1.4. Peptides, Generated With Staphylococcal Protease, V8

Figure 5.5 shows the electrophoresis map of the distribution of peptides in the G25 gel filtration eluate and the corresponding 280 nm absorption trace. The aim of this experiment was to obtain a peptide that covered the C-terminal end, including the suggested tryptophan in position 75, and to acquire independent evidence for the sequence on the basis of the amino acid composition of the peptides. Only the most intense spots were eluted and subjected to amino acid analysis. The peptides obtained appeared to cover the whole sequence. All the glu bonds were completely hydrolysed, apart from the 80-glu-gln-81 bond, which was only partially hydrolysed. Tables 5.5 and 5.6 show, respectively, the amino acid composition of individual peptides, and a summary of other relevant information.

Comments on individual peptides are below. Residue designations starting with a capital letter signify residue identities obtained by sequencing of that particular peptide. Residue designations starting with a lower case letter, signify residue identities obtained by amino acid composition analysis, only. When residues are connected with a hyphen, it signifies that they are in the order which was determined by sequencing of the corresponding peptide or some other peptide from the same region.

V1: This peptide was isolated in a relatively high yield. On HVPE pH 3.5 it runs in at least three different forms, probably reflecting the three electrophoretic variants, -asn-gly- ,-aspα-gly- and -aspβ -gly- (see Th6).

The amino acid composition shows only two isoleucines, whereas the sequencing gives three. This had been previously observed for this sequence part in other peptides. Other integral values are in good agreement with the deduced sequence. Automated sequencing of this peptide was attempted, but it did not get past the asn-gly pair. The proposed sequence: 50-Ile-(ala-asx-ile-ile-asx-asx-gly-lys-asx-ala-met-pro-ala-gly-met-val-asx-pro-ala-lys-ala)-glu-72.

V2: Qualitative and quantitative amino acid, and N-terminal amino acid analyses, were performed. Yellow spot emerged after HVPE pH 6.5 and ninhydrin/collidine staining. The proposed sequence: **32-Gly-(gly-val-gly-pro-ala-leu-ala-asx-val-gly-ser-arg-leu-ser-ala-glx)-glu-49.**

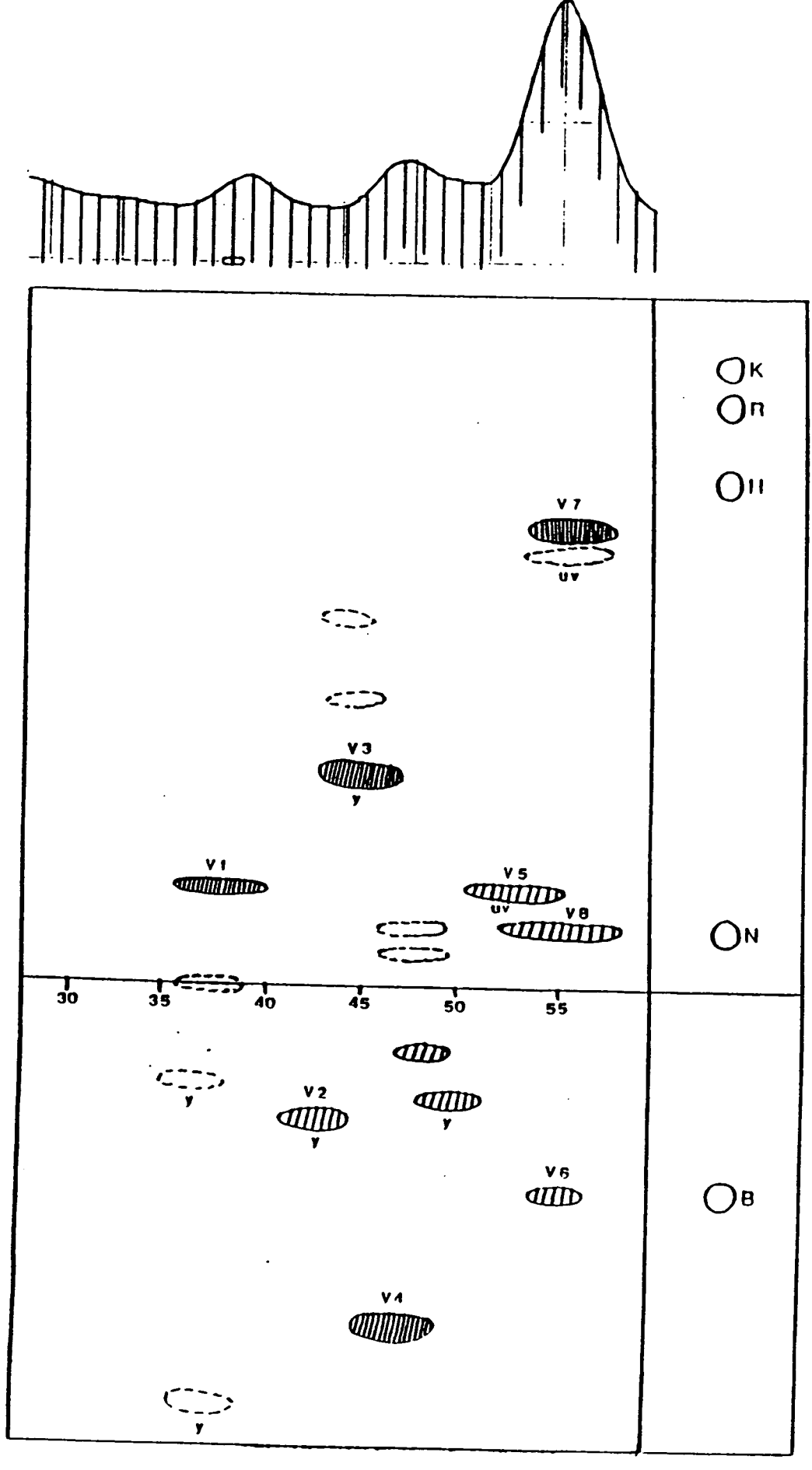


Figure 5.5. HVPE, pH 6.5 mapping of peptides generated with staphylococcal protease, fractionated by G25 gel filtration. The corresponding 280 nm absorption trace of the gel elution is shown on top.

Table 5.5. Amino acid composition of peptides from the P1-c552 apocytochrome generated with staphylococcal protease V8.

Amino acid	V1	V2	V3	V4	V5	V6	V7 & V8
Lys	1.84 (2)	(0)	0.69 (1)	(0)	0.84 (1)	(0)	1.01 (1)
His	(0)	(0)	+++ *(1)	(0)	(0)	(0)	(0)
Arg	(0)	1.00 (1)	(0)	(0)	(0)	(0)	(0)
Asx	5.39 (5)	1.04 (1)	1.09 (1)	2.07 (2)	(0)	(0)	(0)
Thr	(0)	(0)	(0)	(0)	(0)	0.95 (1)	(0)
Ser	(0)	1.85 (2)	1.66 (2)	(0)	(0)	(0)	(0)
Glx	1.05 (1)	2.16 (2)	2.26 (2)	2.24 (2)	2.07 (2)	2.13 (2)	1.00 (1)N
Pro	1.59 (2)	1.35 (1)	(0)	1.87 (2)	(0)	(0)	(0)
Gly	2.05 (2)	4.02 (4)N	2.09 (2)N	(0)	(0)	0.97 (1)N	(0)
Ala	4.96*(5)	3.20 (3)	2.17 (2)	4.16 (4)	3.00 (3)	1.08 (1)	(0)
Val	0.97 (1)	1.95 (2)	0.94 (1)	(0)	0.52 (1)N	(0)	(0)
Met	1.67 (2)	(0)	1.00 (1)	(0)	(0)	(0)	(0)
Ile	2.23*(3)N	(0)	(0)	(0)	0.56 (1)	(0)	(0)
Leu	(0)	1.95 (2)	1.07 (1)	(0)	0.89 (1)	(0)	(0)
Tyr	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Phe	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Cys	(0)	(0)	1.46 (2)	(0)	(0)	(0)	(0)
Trp	(0)	(0)	(0)	(0)	(1)	(0)	(0)
Total	23	18	16	10	10	5	2
Seq. pos.	I50-E72	G32-E49	G16-E31	Q6-E15	G27-E31	V73-K82	G1-E5

Amounts used for composition analysis of peptides, 5-20nm. Values less than 0.1 are omitted.
Values in parenthesis are deduced from the sequence. Comments on individual peptides are in text.
Values marked with an asterix (*) are incongruent with deduced sequence.
N: designates a N-terminal amino acid.
+/++/+++ : qualitative amino acid analysis only.

Table 5.6. General information on peptides generated with staphylococcal protease V8 from cytochrome P1-c552.

Peptide	Purification[a]	V/Vo (G25)	Mobility[b]	Yield[c] nmol %	N-term. a.acid	number a.acids	position
V1	G25 6.5	0.95	0.05	257 17	Ile	23	50-72
V2	G25 6.5 3.5	1.07	-0.35	190 13	Gly	18	32-49
V3	G25 6.5	1.13	0.26	103 7	Gly	16	16-31
V4	G25 6 5	1.17	-0.75	149 10	Ala	10	6 -15
V5	G25 6.5 HPLC	1.31	0.05	69 5	Val	10	73-82
V6	G25 6.5	1.34	-0.46	154 10	Gly	5	1 - 5
V7	G25 6.5	1.36	0.66	125 8	Glx	2	81-82
V8	G25 6.5	1.38	0.00	56 4	Blocked	2	81-82

[a] G25: gel filtration; 6.5: HVPE pH 6.5
[b] mobility in pH 6.5 HVPE relative to lysine (+1)
[c] % : percentage recovery of protein digested (1500 nmol).

V3: This is a haem peptide. Qualitative and quantitative amino acid and N-terminal amino acid analyses were performed. Yellow spot emerged by ninhydrin/collidine staining after HVPE pH 6.5. The proposed sequence: **16- Gly-(met-val-lys-ala-ser-cys-gly-glx-cys-his-gly-glx-asx-leu)-glu-32.**

V4: Qualitative and quantitative amino acid, analysis and N-terminal amino acid analyses were performed. The N-terminal was blocked. The great acidic mobility suggests two aspartates. The proposed sequence: **6-gln-(ala-ala-pro-ala-ala-asx-asx-pro)-glu-15.**

V5: This peptide spot was fluorescent under UV light, after the HVPE pH 6.5 and before ninhydrin/collidine staining. This suggests the presence of tryptophan. The peptide was obtained in a very low yield. It was further purified by reverse phase HPLC. This was a hydrophobic peptide, which at high loading on the c_{18} column eluted in a very broad and irregular peak, but with number of sharp peaks rising from it. At a lower (10 times less) loading, one peak, distinctly larger than the others, was observed. This peak was collected for automated sequencing. The peptide sequenced cleanly, but the tryptophan peak was not recognised by the detector program of the sequenator. This might be due to TFA (trifluoroacetic acid) induced modification of the tryptophan. The large peak observed in this cycle was very close to where the tryptophan peak would have been expected. The retention time of this "amino acid" was 23.1 min and is marked onto figure A2 in appendix A. Table A5 in appendix A shows the yields of amino acids in each cycle of the sequencing run. The peptide gave the sought after overlap with the T2 peptide from the tryptic digest. The proposed sequence: **73-Val-Ile-Ala-Ala-Trp-Leu-Ala-Glu-Gln-Lys-82.**

V6: Yellow emerged spot by ninhydrin/collidine staining after HVPE pH 6.5. Only the amino acid composition and N-terminal amino acid analyses were performed. On the basis of this information, this is clearly the N-terminus peptide. The proposed sequence: **1-Gly-(glx-ala-thr)-glu-5.**

V7 and V8: These peptides had the same amino acid composition, but the V8 peptide was blocked. It is proposed that they are the same and the N-terminal is blocked due to the glutamine cyclisation to pyrollidone

carboxylic acid in V7, resulting in the V8 peptide. This is also evident in the very different mobility of these peptides where the blocked peptide has an extra negative charge. The proposed sequence: **81-Gln-(lys)-82.**

5.2. CYTOCHROME P3&P4-C555

P3-c555 and P4-c555 are different forms of the same cytochrome c555 (chapter 4). To obtain sufficient amount (1600 nmol) of the protein for generation, purification and subsequent sequencing, the P3-c555 and P4-c555 fractions were combined. Figure 5.6 shows the proposal for the order of the obtained peptides in the primary structure of the cytochrome, and the schematic representation of the sequencing results.

5.2.1. N-Terminal Sequence

FPLC purified cytochrome P4-c555 was dehaemed and the apoprotein purified by reverse phase HPLC. A single symmetrical peak was collected for automated sequencing. The protein sequenced cleanly for the first 14 amino acids, after that the sequence became very obscure. The amino acid yields in cycles 15, 16 and 20 were very low, and the most probable residues are suggested. The residues in the parentheses are the most probable ones in these positions. Table A6 in appendix A shows the yields of amino acids in each cycle of the sequencing run. Proposed sequence: **Gln-Thr-Val-Glu-Phe-Asp-Lys-Lys-Asp-Pro-Gly-Tyr-Lys-Ala-(leu)-(glu)-Ala-Ser-Gly-(leu).**

5.2.2. Peptides, Generated by Trypsin Proteolysis

Figure 5.7 shows the electrophoresis map of the distribution of peptides in the G25 gel filtration eluate, and the corresponding 280 nm absorption trace. Tables 5.7 and 5.8 show, respectively, the amino acid composition of the individual peptides, and a summary of other relevant information.

Comments on individual peptides are below. Residue designations starting with a capital letter signify residue identities, obtained by sequencing of that particular peptide. Residue designations starting with a lower case letter signify residue identities obtained by amino acid composition analysis, only. When residues are connected with a hyphen, it

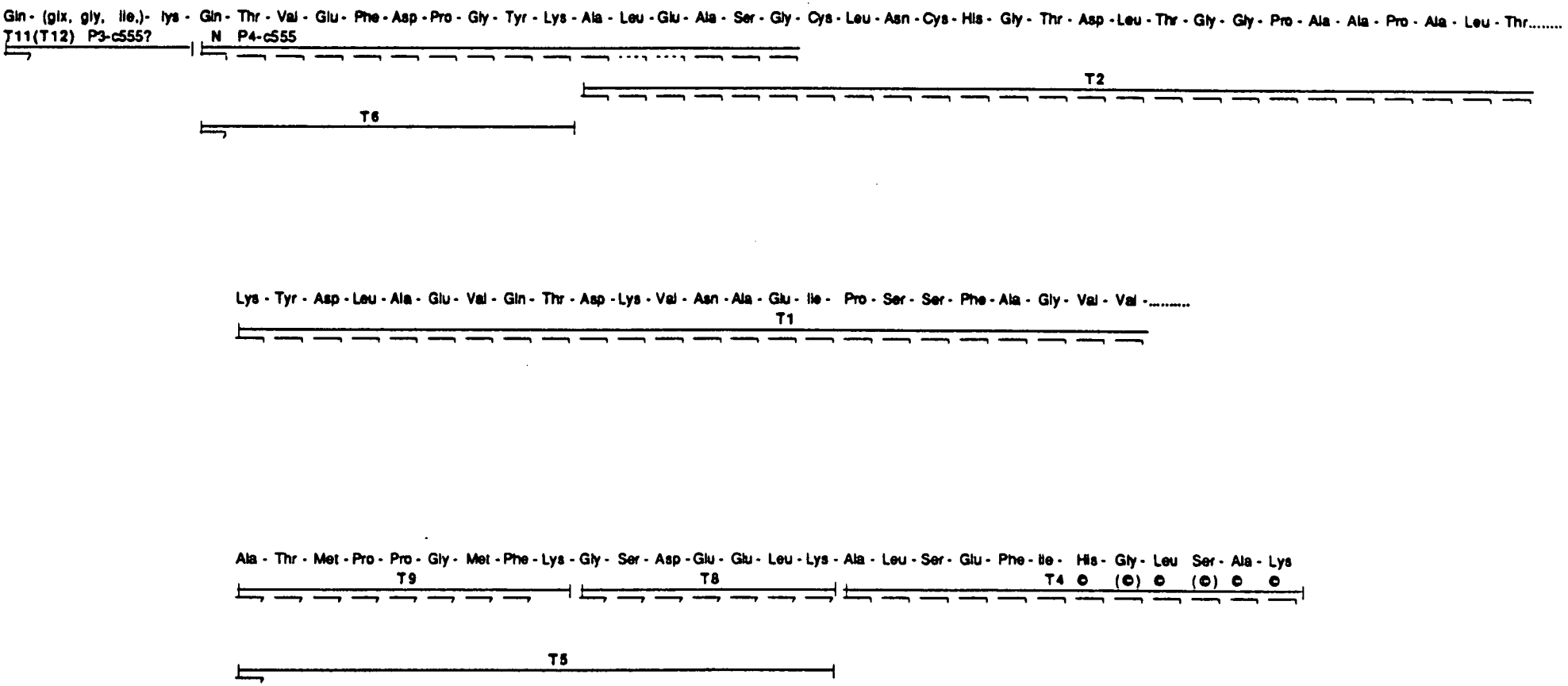


Figure 5.6. Amino acid sequence of *Bacillus azotoformans* cytochrome P3&4-c555.

The peptides are shown in the proposed sequence order for the cytochrome P3&P4-c555 N: N-terminus sequencing. T: Peptides derived by digestion with trypsin. Continuous lines indicate quantitative amino acid analyses. Under peptide lines, ⟶ indicates end groups and subsequent residues revealed by phenyl isothiocyanated degradation and identified by the dansyl method, ⟶ indicates residues obtained by automated sequencing. Broken lines indicate substandard results or a residue unrecognised by the sequenators program. ©: residues indicated by carboxipeptidase digestion and identified by analysis on a Beckman 120C amino acid analyser. ..... , indicates incomplete sequencing of protein or peptide.

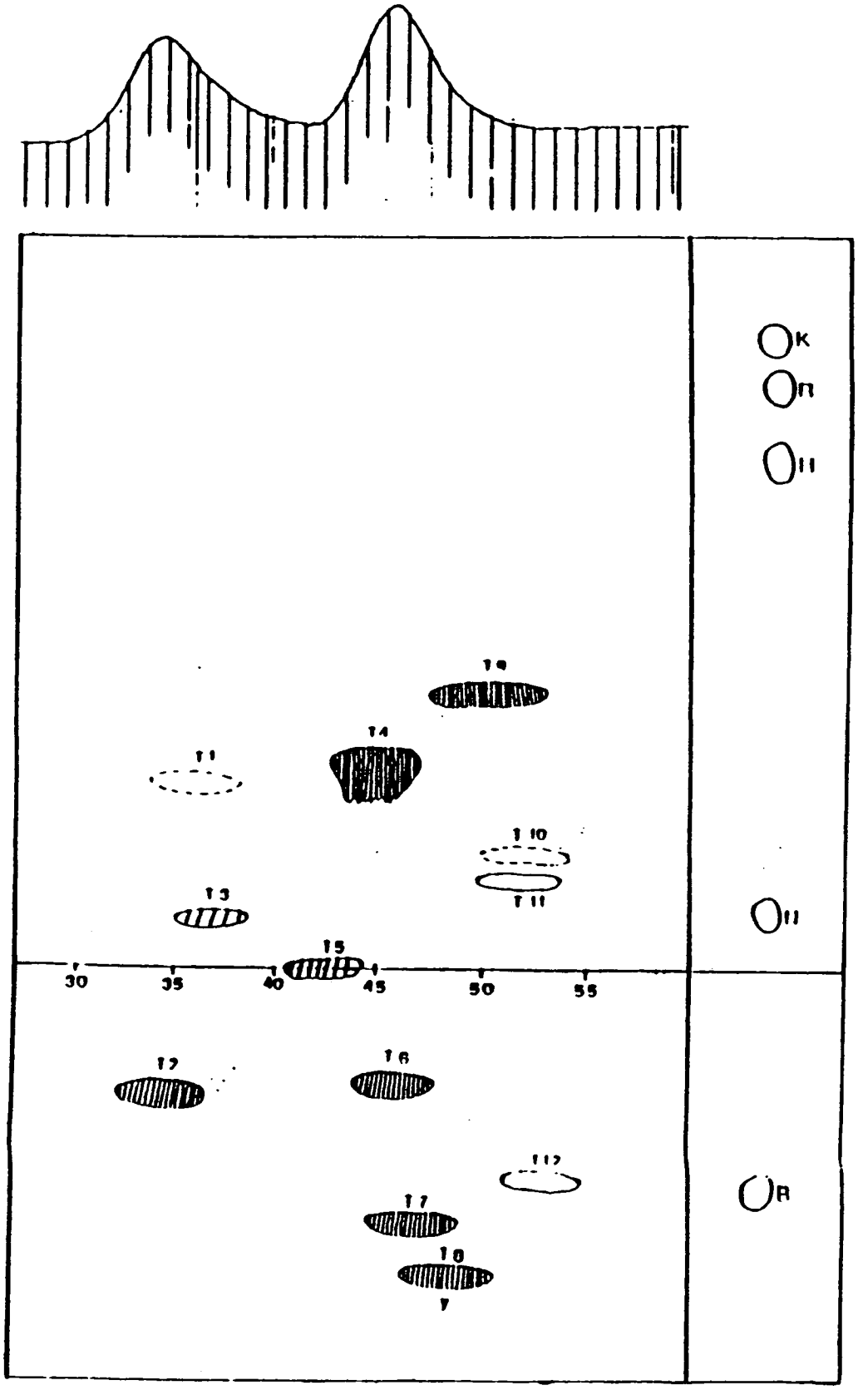


Figure 5.7. HVPE, pH 6.5 mapping of tryptic peptides fractionated by G25 gel filtration. The corresponding 280 nm absorption trace of the gel elution is shown on top.

Table 5.7. Amino acid composition of peptides from the P3&P4-c555 apocytochrome generated with trypsin.

Amino acid	T1	T2	T3	T4	T5	T6 (& T7)	T8	T9	T10	T11& T12
Lys	3.00 (3)N	(0)	0.96 (1)	0.94 (1)	1.93 (2)	2.00 (2)	0.96 (1)	0.96 (1)	++	0.89 (1)
His	(0)	0.67 (1)	0.37 (1)	0.57 (1)	(0)	(0)	(0)	(0)		(0)
Arg	(0)	0.74 (1)	0.45 (1)	(0)	(0)	(0)	(0)	(0)		(0)
Asx	3.25 (3)	3.68 (4)	3.18 (3)	0.15 (0)	1.22 (1)	2.21 (2)	1.05 (1)	(0)		(0)
Thr	1.80 (2)	4.37 (6)	3.09 (3)	(0)	1.05 (1)	1.01 (1)	(0)	0.94 (1)		(0)
Ser	2.85 (3)	2.03 (2)	2.02 (2)	1.63 (2)	1.12 (1)	(0)	1.00 (1)	(0)		2.08 (2)
Glx	4.02 (4)	3.80 (4)	3.61 (4)	0.99 (1)	2.26 (2)	3.47 (3)N	2.26 (2)	(0)	++	(0)
Pro	2.24 (2)	1.27 (2)	1.89 (2)	0.23 (0)	1.91 (2)	0.66 (1)	(0)	1.84 (2)	(+)	1.14 (1)N
Gly	2.53 (3)	5.67 (6)	4.32 (4)	1.06 (1)	2.41 (2)	1.32 (1)	0.97 (1)N	1.07 (1)	+	(0)
Ala	3.45 (3)	8.12 (8)N	5.14 (5)N	1.80 (2)N	1.04 (1)N	(0)	(0)	0.89 (1)N	++N	(0)
Val	3.36 (3)	0.92 (1)	1.20 (1)	(0)	(0)	1.04 (1)	(0)	(0)		(0)
Met	0.45 (1)	(0)	0.35 (1)	0.11 (0)	1.36 (2)	(0)	(0)	1.69 (2)	+	(0)
Ile	1.64 (2)	1,80 (2)	1.47 (2)	0.73 (1)	1.00 (1)	(0)	(0)	(0)		0.97 (1)
Leu	1.69 (2)	4.30 (4)	3.05 (3)	1.69 (2)	(0)	(0)	1.03 (1)	(0)	++	(0)
Tyr	0.94 (1)	(0)	0.35 (1)	(0)	(0)	0.67 (1)	(0)	(0)		(0)
Phe	1.31 (1)	(0)	0.84 (1)	0.89 (1)	0.99 (1)	1.07 (1)	(0)	1.12 (1)	(+)	(0)
Cys	(0)	2.09 (2)	0.33 (2)	(0)	(0)	(0)	(0)	(0)		(0)
Total	33	43	37	12	16	13	7	9	?	5

Amounts used for Beckman 120C analysis: 5-20 nmol. Values less than 0.1 are omitted.
Values in parenthesis are deduced from the sequence. Comments on individual peptides are in text. Values marked with an asterix (*) are incongruent with deduced sequence.
N: designates a N-terminal amino acid.
+/++/+++ : qualitative amino acid analysis only.

Table 5.8. General information on the tryptic peptides from P3&P4-c5552.

Peptide	Purification[a]	V/Vo (G25)	Mobility[b]	Yield[c] nmol	%	N-term. a.acid	number a.acids
T1	G25 6.5 HPLC	0.90	0.80	70	4	Lys	33?
T2	G25 6.5 HPLC	0.87	-0.33	134	8	Ala	43?
T3	G25 6.5	0.91	-0.08	75	5	Ala	37?
T4	G25 6 5	1.12	0.24	121	8	Ala	11
T5	G25 6.5 HPLC	1.06	-0.12	143	9	Ala	16
T6	G25 6.5	1.14	-0.31	102	6	Gln	12
T7	G25 6.5	1.18	-0.51	154	10	Gln	12
T8	G25 6.5	1.20	-0.57	110	9	Gly	7
T9	G25 6.5	1.28	0.36	258	16	Ala	9
T10	G25 6.5 3.5	1.30	0.07	low		Ala	?
T11	G25 6.5 3.5	1.29	0.02	57	4	Gly	5
T12	G25 6.5	1.31	-0.43	118	7	Blocked	5

[a] G25: gel filtration; 6.5: pH HVPE.
[b] mobility in pH 6.5 HVPE relative to aspartate (-1).
[c] % : percentage recovery of protein digested (1600 nmol).

signifies that they are in the order which was determined by sequencing from the corresponding peptide or some other peptide from the same region. Unconnected residues signify amino acid composition of peptides where residue order is unknown.

T1: This peptide was isolated in a low yield. The N-terminal analysis did not give any clear N-terminal amino acid, but there was the slightest hint that it might be lysine. This was, subsequently, confirmed by automated sequencing. The peptide showed UV fluorescence before it was stained with ninhydrin/collidine after HVPE pH 6.5, which suggests the presence of tryptophan in the peptide. The hydrolysis for the quantitative amino acid analysis was done with MESA (see chapter 2), but no tryptophan was detected. According to quantitative amino acid analysis this is a large peptide, about 33-35 residue long. The exact number must, however, remain uncertain, as the integral values deviate to some degree from whole integer values. The peptide was further purified by HPLC and subjected to automated sequencing. The first twenty four residues sequenced cleanly. Table A7 in appendix A shows the yields of amino acids in each cycle of the sequencing run. The proposed sequence: Lys-Tyr-Asp-Leu-Ala-Glu-Val-Gln-Thr-Asp-Lys-Val-Asn-Ala-Glu-Ile-Pro-Ser-Ser-Phe-Ala-Gly-Val-Val-(pro pro glx thr ser ile leu gly gly met)-lys.

T2 and T3: These are the haem site peptides. According to the quantitative amino acid composition analysis, they are 43 (T2) and 37 (T3) residues long. The peptides have identical N-terminal amino acids, and they are similar in amino acid composition. However, they differ in aspects which is difficult to reconcile with the fact that they are derived from the same region of the protein. The larger peptide, T2, which elutes a bit earlier in the gel filtration purification step, has one arginine and no lysine. The smaller peptide, T3, has one of each. The electrophoretic mobilities obtained by the first HVPE pH 6.5 analysis also confirmed this difference of one positively charged amino acid. As the peptides were obtained by trypsin proteolysis, one would have expected the larger peptide, T2, to have the two basic amino acids, lysine and arginine, as the precursor of the the smaller peptide, but not the other way around.

When samples of each peptide were subjected to HVPE pH 3.5 analysis, they had apparently precipitated and they stained badly. It was also evident that both were contaminated and the T2 peptide to a greater extent.

Both were seemingly contaminated with hydrophobic amino acids or peptides, but in addition the HVPE map for the T2 peptide (the larger peptide), revealed a ´satellite´ peptide, with a greater mobility. Surprisingly, this contamination also showed up in a parallel analysis experiment by HVPE pH 6.5 which would have been expected to give only one spot, identical to the one obtained in the preceding preparative run (also HVPE pH 6.5). It may be concluded that the peptides are very likely the same, apart from the one additional residue, lysine, in the T3 peptide, and that the difference in composition is because of contamination.

The HVPE pH 3.5 analysis was performed soon after the elution from the preparative HVPE pH 6.5. The preparations were then frozen for storage. On thawing, it became evident that both had come out of solution. When the pH 3.5 and pH 6.5 HVPE analyses experiments were repeated, neither peptide showed up by ninhydrin/collidine staining in either run. A further purification by HVPE pH 3.5 or repeated HVPE pH 6.5, was deemed futile.

The T2 peptide was selected for automated sequencing. It was first spun down and the supernatant set aside. It was reasoned that the precipitate consisted mainly of the large haem peptide and that the contaminating peptide would be still in solution. The precipitate was dissolved in 1% TFA (trifluoroacetic acid) and subjected to reverse phase hydrophobic chromatography on a High bore HPLC (column C_{18}). The peptide eluted close to the end of the run. The peak was broad and unsymmetrical. The reversed hydrophobic chromatography was repeated for the fraction, containing the highest part of the peak. In this run the peptide had about half the retention time of the previous run, which indicated that the solubility of the peptide had become greater. The fraction containing the peak was dried down and sequenced.

The first 25 residues sequenced cleanly, after which the sequence became unreadable. A sharp fall in the amino acid signal was observed in the subsequent cycles, from approximately 300 pmol down to 3 pmol. The sequence is shown below and also the proposed composition of the part still to be sequenced. It is based on the amino acids composition of the smaller haem peptide T3, on the basis of the arguments given above. The aromatic residues, which only show up in this peptide are omitted as probably contaminants of this peptide. The C-terminal is proposed to end in arg-lys.

A possible longer sequence may be obtained with higher initial loading on the sequenator (1000pmol). If the digest and purification is repeated, it

is advisable to process and sequence the peptides without an intermediate period of freezing.

Table A8 in appendix A shows the yields of amino acids in each cycle of the sequencng run. The proposed sequence: **Ala-Leu-Glu-Ala-Ser-Gly-Cys-Leu-Asn-Cys-His-Gly-Thr-Asp-Leu-Thr-Gly-Gly-Pro-Ala-Ala-Pro-Ala-Leu-Thr-(asx-ser-val-met-ile-ile)-(arg-lys)**

T4: This peptide streaks to some extent in the HVPE pH 6.5 electrophoresis, presumably because of partial precipitation. The peptide was completely sequenced by automated sequencing. Table A9 in appendix A shows the yields of amino acids in each cycle of the sequencing run. The proposed sequence: **Ala-Leu-Ser-Glu-Phe-Ile-His-Gly-Leu-Ser-Ala-Lys.**

T5: On the basis of quantitative amino acid- and N-terminal analyses it was concluded, that this peptide derived from an incomplete cleavage and therefore it had two lysines. Two smaller peptides, T8 and T9 which were sequenced, are proposed to cover this same region. On the basis of that information, the sequence is proposed to be. **ala-thr-met-pro-pro-gly-met-phe-lys-gly-ser-asx-glx-glx-leu-lys.**

T6 and T7: These peptides have a similar amino acid composition. They differ in the number of glx (glutamine/glutamate), with T7 having one less. The T7 was blocked at the N-terminus so that a glutamine at the N-terminus of T7 may have cyclisised to form pyrollidone carboxylic acid. This is in accord with its greater acidic mobility. The amino acid composition in table 5.7 is that of **T6**. The composition accords with the N-terminal sequence of the protein (dehaemed **P3&P4-c555**). It was not sequenced. The proposed sequence: **gln-(thr-val-glx-phe-lys-glx-asx-pro-gly-tyr)-lys**

T8: On the basis of the great acidic electrophoretic mobility on HVPE pH 6.5, it was concluded that the peptides contained two glutamates and one aspartate (see figure 5.7). The peptide was sequenced, manually. The proposed sequence: **Gly-Ser-Asx-Glx-Glx-Leu-Lys.**

T9: This peptide was sequenced manually. The quantitative amino acid analysis gave good integral values, but as the yield of the peptide was relatively low, only seven cycles were done. The remaining amino acid

according to the amino acid composition is deduced to be lysine. This was confirmed by automated sequencing (table A10 in appendix A). This is to be expected in peptides obtained by tryptic digest. The proposed sequence: **Ala-Thr-Met-Pro-Pro-Gly-Met-Phe-(lys).**

T10: This peptide was obtained in a very low yield and only qualitative amino acid composition and N-terminal analyses were performed.

T11 & T12: These peptides have the same amino acid composition, but they differ considerably in mobility. The N-terminal was blocked in the peptide which was obtained in a greater yield, T11, but the N-terminal amino acid was glutamine/glutamate in the T12 peptide. The difference in mobility, between these peptides, can be explained if glutamine is the N-terminal amino acid and that it had undergone cyclisation to pyrollidone carboxylic acid in the T11 peptide. It would then explain the blockage of the N-terminal of the T11 peptide. The proposed sequence: **Gln-(glx-gly-ile)-(lys).**

5.2.3. C-Terminal Study on Cytochrome P3&P4-c555

Approximately 20 nmol of cytochrome **c555** were digested with 100 μ g carboxypeptidaseA for 4 hours. The solution was dried down and analysed directly on the amino acid analyser. The following amino acids were released.

	Control nmol	P3-C555 nmol	P4-C555 nmol
Ala	3.1	3.4	3.2
Gly	3. 9	2.7	4.1
Leu		2.7	2.8
Ser		2.5	3.2
His		2.5	2.7
Lys		4.3	4.9

P3-c555 and P4-c555 are different chromatographic isolates (FPLC) of the same cytochrome, c555. It is apparent that the proteins are not very susceptible to carboxypeptidase digestion and the amounts of released amino acids are near the limit of sensitivity for the amino analyser. Still, some digestion seems to have occurred and even if the control showed approximately the same amounts of the amino acids, glycine and alanine,

the pattern of amino acid release is consistent with the T8 peptide composition. No other peptide, for various reasons such as being already confidently positioned or not having the right amino acid composition, fits this pattern. If correctly interpreted these results indicate that the difference in size and chromatographic properties is due to a difference in the N-terminal part of the cytochrome c555.

5.3. Cytochrome P2-C551

Only a small amount of this protein was obtained. It was purified according to the procedure described in chapter 4. The last step was FPLC chromatography on Mono-Q, This cytochrome c was dehaemed, and then subjected to reverse phase hydrophobic chromatography (HPLC). It eluted as a doublet of two sharp peaks. One of these was collected and subjected to automated sequencing. The protein sequenced cleanly for the first 10 amino acids. After cycle 10, the sequence seems blocked. Table A11 in appendix A shows the yields of amino acids in each cycle of the sequencing run. Proposed sequence:

Glu-Pro-Ala-Asp-Asn-Asn-Ala-Gly-Asp-Gly

5.4. Cytochrome P5-C552

This protein was also obtained in a low yield. It was purified according to the P2-c551, as has been described in chapter 4. It was purified further by reverse phase HPLC and subjected to automated sequencing. Table A12 in appendix A shows the yields of amino acids in each cycle of the sequencing run. The sequence obtained was 37 residue long. Residues in parentheses were obtained in very low yields. Proposed sequence:

Glu-Gln-Thr-Thr-Glu-Thr-Ala-Gln-Pro-Asn-Leu-Ala-Glu-

(his)-Thr-Asp-Asn-Ile-Leu-Val-Gln-Lys-Gly-(cys)-Ile-Ala-

(cys)-(his)-Ala-Val-Ser-Pro-Ile-Gly-Ala-Lys-Gly

CHAPTER 6
6. ANALYSES AND DISCUSSION

6.1. THE GENUS *BACILLUS*

The genus *Bacillus* has played a major role in the development of microbiology since it was created in 1872 by Ferdinand Cohn. While no single phenotypic trait or characteristic is unique to the genus, of those which collectively describe or delineate *Bacillus*, many are unique, or nearly so, to the Gram positive phylum. This includes the positive response to the Gram stain, which was developed by Christian Gram in 1884. It reflects the characteristic structure and composition of the cell wall of Gram positive bacteria. Compared to Gram negative bacteria they have a thicker wall, which is made up of a greater amount of peptidoglycan and more extensively crosslinked. They have a lesser percentage of lipids and no outer membrane. Another characteristic, which is a distinct property of the Gram positive bacteria, is their ability to form spores. In 1877 Cohn described the heat resistant spores of *Bacillus subtilis*, and consequently put an end to the theory of spontaneous generation. Sporulation is first incorporated in a classification key of bacteria in 1884 by DeBarry and VanThieghem. It is now recognised as one of the most distinctive characteristics of the phylum, but it is also now generally accepted that many Gram positive strains and species have lost this ability. In 1907 aerobic growth was introduced by Flugge as an important characteristic to distinguish *Bacillus* from remaining rod shaped endospore forming bacteria (Gordon,1981).

It is clear from looking at the history of the genus as recounted by Gordon (1981), that the consensus of characteristics used for the definition of the genus *Bacillus* has changed, and an increasing number of more delineating traits have been added with time and technical progress. However, by and large, the simple definition in Bergey´s manual, given below in a shortened version, suffices for most purposes to identify bacteria of the genus.

The genus *Bacillus* (Claus and Berkeley., 1986): *Rod shaped, straight; endospores, very resistant to adverse conditions; one spore per cell; sporulation not repressed by exposure to air; Gram positive; flagella peritrichous; aerobic or facultatively anaerobic; catalase formed by most species; chemoorganotrophs; the cell wall peptidoglycan of most species belongs to the directly cross linked meso-diaminopimelic acid type.*

The description encompasses bacteria of very different origins and adaptations, with representatives living at the extremes of both temperature and pH. A wide diversity of properties is observed in the genus. This is reflected in the wide range of nutritional requirements and growth conditions spanned by its members and is emphasised by the G+C% spectrum, 32-68%, observed, which is far wider than is generally considered reasonable within a single genus. (Norris *et al.* 1981). The genus *Bacillus* is therefore an interesting subject for purely taxonomic reasons and one to which the techniques of molecular biology can be expected to have great importance.

The general metabolic character of the genus *Bacillus* has significance for the elucidation and interpretation of the evolutionary route of Gram positive bacteria, for evaluating their placement in the eubacterial tree and status in the history of metabolic pathways.

From the early age of microbiology the traditional opposites to Gram positive bacteria have been the purple bacteria. They are in essence the Gram negative bacteria. Consequently the information gained about the metabolic apparatus of the Gram positive bacteria will always be interesting, viewed from this vantage point. Perhaps surprising similarities will be found as well as differences reinforcing the dichotomy.

6.1.1. Cytochromes C and Other Electron Transport Proteins from The Genus *Bacillus*

Bacillus subtilis was one of the first micro-organism in which cytochrome c was detected (Keilin, 1925). During the six decades since this event, physiological studies have mostly been confined to spectrophotometrical analyses of the electron transport components under different growth conditions. Surprisingly little information was gathered on the specific functional roles, variety and distribution of cytochromes c and other components of the electron transport pathways

in the genus *Bacillus*. The studies were few and far apart, and limited in scope in comparison with the extensive work on the electron transport pathway in Gram negative bacteria. Sequence information has also been scarce.

There are indications that the picture may be rapidly changing. Recently important studies on electron transport proteins in various species of *Bacillus* have been published. An overview of the main results obtained is given below and also relevant information obtained in the present study.

6.1.1.1. Spectral analyses

In his early studies, Keilin (1925) identified a-type and b-type cytochromes as well as cytochrome c, in *B. subtilis*. The presence of these cytochrome types has since been confirmed in other species and genera of the Gram positive phylum.

The predominant pattern found in heterotrophic Gram positive bacteria, such as the genus *Bacillus* of the low mol% G+C phylogenetic group (the *Lactobacillus, Clostridium, Bacillus* lineage), has the composition bcaa3o, which indicates that they are a homogenous group. When cells of the heterotrophic *Arthrobacter globiformis* of the high mol% G+C Actinomycete phylogenetic group, become oxygen limited, haem d is added and the pattern becomes bcaa3od (Jones, 1986).

Cytochrome c is often absent from facultative anaerobic bacteria and lactic bacteria appear to contain only cytochrome b. *Propionibacteria* have the pattern bda1 (Jones, 1986). *Clostridium* usually lacks cytochromes (Jones, 1986), but there are reports of cytochromes b in some chemotrophic species such as *Clostridium formicoaceticum* (Thauer et al,. 1977). The only photosynthetic member of the phylum, *Heliobacterium chlorum*, appears to have b and c type cytochromes (Fuller *et al.*, 1985). Its Gram negative analogues, *Selenomonas* and *Sporomusa* have b type cytochromes whereas *Megasphera* has none (DeVries, 1973, Stackebrandt, 1985).

6.1.1.2. Sequence information on cytochromes c and other components of the electron transport chain in the genus *Bacillus*

Only cytochromes c of Class I, as defined by Ambler (1977), have been detected in the genus of *Bacillus* and all but one of those, which have been isolated and sequenced from the genus, are of the same structural type. The sequences are; a complete haem domain sequence of cytochrome c550 and another partial and tentative sequence of cytochrome c551 from *B. licheniformis* (Woolley, 1984; van Beeumen, personal communication), one complete sequence of cytochrome c550 from *B. subtilis* (VonWachenfeldt and Hederstedt, 1990), one complete sequence of cytochrome c551 from the thermophilic *Bacillus* strain PS3 (Fujiwara *et al.*, 1993) and a partial sequence from *Bacillus halodenitrificans*, c550 (Saraiva *et al.*, 1992). **This structural type will be called BacI cytochromes in the following discussion.**

The odd sequence that does not belong to the BacI sequence group, is a component of the caa3 cytochrome oxidase complex in *Bacillus* PS3 (Ishizuka *et al.*, 1990). It differs from them in major aspects. (see sections 6.3.1 to 6.3.4).

In this thesis is presented sequence information on the haem domains of four different apparently membrane bound cytochromes from *Bacillus azotoformans*, two of which may be of different structural type than those previously sequenced from the genus.

Sequence information on other electron transport proteins includes the succinic dehydrogenase gene from *Bacillus subtilis*, which is a b-558 cytochrome. It is deduced to have the molecular weight of 22 kD. This cytochrome has approximately twice the size of the cytochrome b subunit of the succinic dehydrogenase in *E. coli*. The separate cytochrome b subunits in *E. coli* may be fused into a single gene product in *Bacillus*. There is, however, no obvious sequence similarity between the *Bacillus* and *E. coli* b-type cytochromes (Frieden *et al.*, 1990, cited in VonWachenfeldt and Hedterstedt, 1990).

Recently the genes for the subunits of aa3 and caa3 oxidases have been sequenced from *Bacillus subtilis* and the thermophilic *Bacillus* species PS3 (Santana et al, 1992, Ishizuka *et al.*,1990). They are homologous to aa3 oxidases from *Paracoccus denitrificans*, *Thermus thermophilus* and to mitochondrial aa3, and perhaps more surprisingly to the bo-oxidase from *E. coli* (Ishizuka *et al.*, 1990).

6.1.1.3. Physiological studies on the role of cytochromes c and other electron transport proteins in bacteria of the genus *Bacillus*

A composite schematic picture of the main components of the electron transport chain, and sequential order of electron transport events, is represented in Figure 6.1. It is based on results from various recent studies. These results are discussed below.

Aerobic conditions

Miki and Okunuki(1969) isolated and purified two different cytochromes from an aerobically grown strain of *Bacillus subtilis*. One of those had similar spectral properties as the sequenced BacI cytochromes, with an alpha absorption peak at 550 nm. The other cytochrome c had quite distinct spectral properties, including a split alpha absorption peak, with the main peak at 554 nm and the shoulder at 549 nm. On the basis of physicochemical and spectral response to different electron donors (NADH, etc.) and reactivity with different oxidases, it was proposed that the former was similar to the mitochondrial c2-type cytochromes and the latter (c554) had physiological significance as a bacterial cytochrome c1 forming a complex with a b-type cytochrome.

VonWachenfeldt *et al.* (1990) isolated the structural gene for cytochrome c550 from *Bacillus subtilis*, which was expressed under aerobic conditions. By deletion analysis he demonstrated that the product of the gene was not essential for aerobic growth as another cytochrome c, with an identical α-peak absorption at 550 nm, was revealed or was expressed instead.

Sone and co-workers have studied extensively the electron transport pathway of a thermophilic *Bacillus* species, designated PS3. They detected three different cytochromes c in this species, two of which were expressed under aerobic growth conditions (Sone and Yanagita, 1982, Sone *et al.*,1988; Ishizuka *et al.*, 1990; Kuoth and Sone, 1988). The Sone group presents physiological evidence that no other cytochrome c is required in the main respiration chain of this species under vigorously aerated conditions. The third cytochrome c was only observed under air-limited conditions (Sone *et al.*, 1983,1989)

One of the aerobically expressed cytochromes c in *Bacillus* PS3 appears to be fused to one of the subunits (subunit COII) of the cytochrome c oxidase aa3 (Ishizuka *et al.*, 1990). The cytochrome c cannot be separated from the COII subunit on SDS gels and the oxidase complex is accordingly

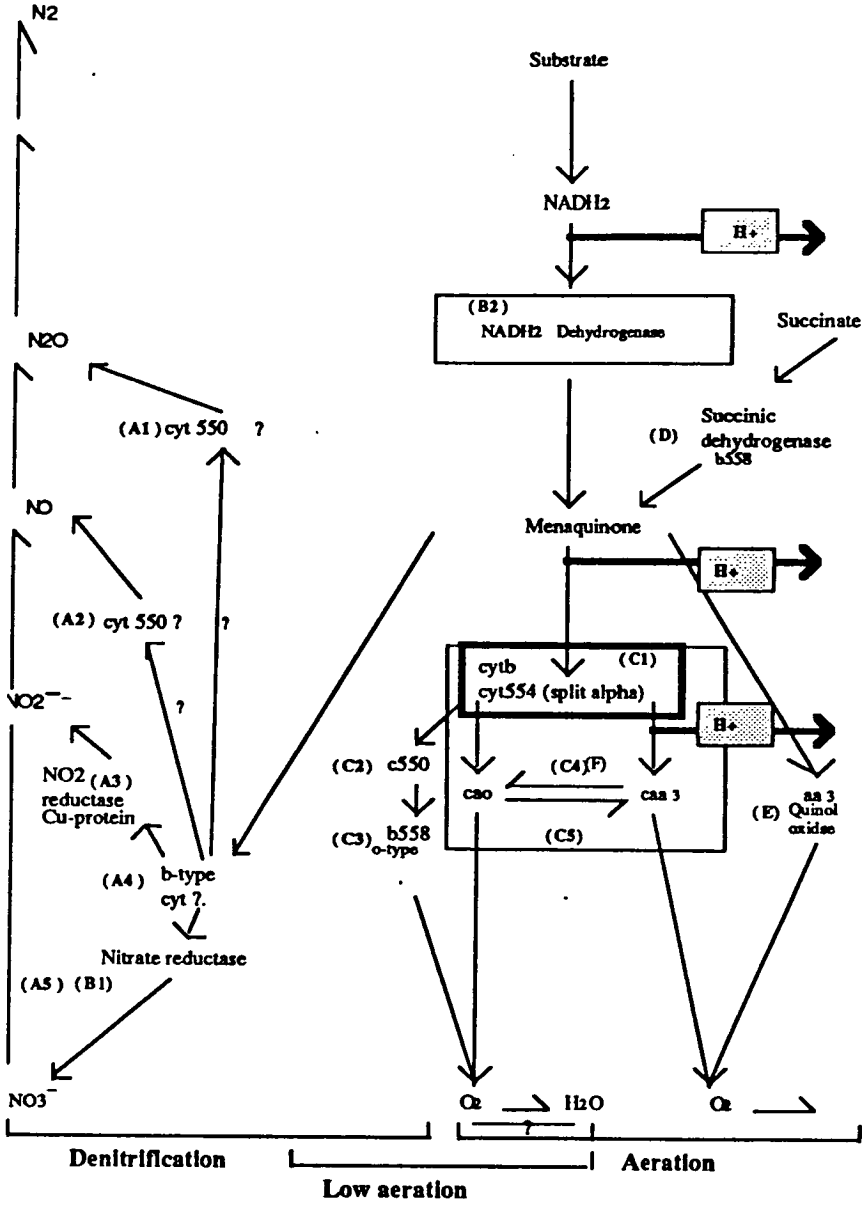


Figure 6.1. A schematic composite figure of the main components in the aerobic and denitrifying electron transport chains in the genus of *Bacillus*, and the sequential order of electron transport events.

Main references: Components of the denitrification pathway, (A1), (A2), (A3), (A4), (A5), Saraiva, 1992; Denariaz, 1991; (B1), Jones 1985; NADH dehydrogenase, (B2), Jones 1985; menaquinone, (B3), Jones 1985; menaquinol reductase complex, (C1), Sone et al., 1982, 1988; Ishizuka et al.,1990; Kuoth and Sone, 1988; cytochrome c550, (C2), Fujiwara et al. 1993; o-type oxidase, (C3), Sone et al 1989; cytochrome c oxidase caa3 (cao) (C4), Sone et al., 1982; Fujiwara and Sone, 1991; cytochrome c oxidase caa3/ menaquinol reductase complex, (C5) Fujiwara and Sone, 1991; Succinic dehydrogenase (D), Frieden et al.,1990; aa3 quinol oxidase (E), Santana et al., 1991.

designated caa3. The oxidase shows a clear proton pump activity (Sone and Yanagita, 1982) and is repressed under severely air-limited growth conditions.

The second cytochrome c, detected in *Bacillus* PS3 under aerobic growth conditions, has a split alpha absorption peak at 554 nm (Sone *et al.* 1983). This cytochrome has very similar spectral properties and is likely to be homologous to the c554 cytochromes c in *B. licheniformis* (Woolley, 1984) and *Bacillus subtilis* (Miki and Okunuki, 1969). It is also found in *B. azotoformans* (P3&P4-c555) both under aerobic and anaerobic denitrifying conditions (this work).

Sone *et al.* (Sone *et al.*, 1983; Kuoth and Sone, 1988) observed that the c554 cytochrome formed a quinol reductase complex with a b-type cytochrome. They propose that it is a c1- or an f-type cytochrome, which are parts of similar quinol reductases in mitochondria and purple bacteria, and algae and cyanobacteria, respectively. They also reported that the similarity of the *Bacillus* PS3 complex is greater to the b6f complex of cyanobacteria than to the one of mitochondria and purple bacteria. Similar conclusions were also reached by Miki and Okunuki (1969) But, while the split alpha cytochrome may be the physiological counterpart to the c1 or the f cytochromes, sequence evidence presented in this thesis shows that the split alpha cytochrome c in *Bacillus azotoformans* has quite different sequence characteristics (see section 6.3.3.). The bc554 quinol reductase complex and the caa3 oxidase form a supercomplex in *Bacillus* PS3 (Sone *et al.*, 1987).

De Vrij (1983) has isolated and purified to homogeneity an aa3 oxidase from aerobically grown *B. subtilis.* In contrast to the results of Sone *et al.* noted above, he did not detect any cytochrome c in the purified fractions. This may be explained by the recent results of Santana *et al.* (1992) and Lauraeus et al, (1991), which have presented genetic and physiological evidence for the presence of two related oxidases in aerobically grown *B. subtilis.* One is an aa3 oxidase and the other is a caa3 cytochrome c oxidase. These oxidases catalyse, respectively, quinol and cytochrome c oxidation,

Low aeration and anaerobic denitrifying conditions

Induction of cytochromes c under conditions of low aeration has been observed in Gram negative bacteria and especially under denitrifying conditions. In some species, it may partly be explained by the occurrence of a nitrite reductase with a domain containing a c-type haem, but soluble cytochromes c are also induced or their expression increased. Induction or expression of one particular subclass of cytochromes c, subclass ID, has been correlated with denitrifying activity (Moore and Pettigrew, 1990). The effect of denitrifying conditions on cytochrome c production has been examined in *Bacillus* species, but with contradictory results. Schulp and Stouthamer (1970) observed a decrease in total cytochrome c content in *B. licheniformis*. In contrast Woolley (1984) observed an increased expression of cytochromes c under denitrifying conditions in a different strain of *B. licheniformis*. In the latter study the yield of split alpha c554 cytochrome was mainly increased. This is surprising because of the proposed essential role of this cytochrome in the aerobic electron transport chain (Sone *et al.*, 1988). As a part of the b6c554 complex it would be expected to be an essential and non limiting component of the aerobic electronic pathway and either constitutively expressed or repressed under conditions of low aeration. The observed increase in yield under denitrifying conditions may therefore be an artefact of the isolation.

Cytochromes c have been isolated from *Bacillus* strains grown under low aeration or denitrification conditions. A cytochrome c550 (subclass BacI) was isolated and partially sequenced from *Bacillus halodenitrificans* (Saraiva *et al.*, 1992) and cytochrome c551 (subclass BacI) from *Bacillus* PS3 was isolated and sequenced from cultures grown under conditions of low aeration (Fujiwara *et al.*, 1993). These cytochromes are apparently induced under these conditions. Fujiwara *et al.* (1993) propose that the cytochrome c551 mediates the electron flow between the cytochrome a bc554 quinol reductase complex and a b558 terminal oxidase, in *Bacillus* PS3. It is noteworthy that these induced cytochromes, as indeed all the BacI sequences, are more similar to the ID subclass cytochrome than to any others (see section 6.3.4). The ID subclass cytochromes, as noted above, are often induced under conditions of anaerobic denitrification in purple bacteria (Moore and Pettigrew, 1990).

The caa3 cytochrome c oxidase is repressed in the *Bacillus* PS3 strain under slightly air limited conditions, or rather, the selective repression of the haem a3 is observed under these conditions. A cao oxidase is formed which is derived from the former by haem substitution (Sone and

Fujiwara, 1991). The latter oxidase does not have proton pumping activity, but it has a higher affinity for O_2.

An aco cytochrome oxidase is constitutive in the alkalophilic *Bacillus* strain YN-2000 (Qureshi *et al.*, 1990). No evidence was found of a haem a3. It is proposed that there is some difficulty in the transport of O_2 through the cytoplasmic membrane in this species.

An o-type oxidase, (b558), appears to be induced and the cao oxidase repressed under severely air-limited growth conditions in *Bacillus* PS3 (Sone *et al.*, 1989). This has also been observed, under anaerobic denitrifying conditions in different mesophilic *Bacillus* species (Pichinoty, 1978).

It is noteworthy that the nitrite reductase in *Bacillus halodenitrificans* is not a cd1 cytochrome, but a copper protein (Denariaz *et al.*, 1991). Both types are found in purple bacteria but sequence information on the copper protein in *Bacillus* is lacking, so that whether they are homologous or not, is left to be established. Physicochemical properties are, however, similar.

The nitrate reductase in *B. licheniformis* is a two subunit enzyme that contains molybdenum and an iron sulphur centre. It is similar in physicochemical properties to nitrate reductases found in Gram negative bacteria (Van´t Riet *et al.*, 1979).

6.1.1.4. *Bacillus azotoformans*

Figure 6.2 is a schematic presentation of the SDS-PAGE profile of cytochromes from *Bacillus azotoformans* grown under different conditions, aerobically, low aeration and anaerobically by denitrification. Two cytochromes were observed, by SDS-PAGE analyses, to be induced under aerobic conditions. One of them, a high molecular cytochrome c (>100 kD, band 1 in figure 6.2), was repressed under air-limited growth conditions without nitrate, as well as under anaerobic denitrifying conditions. This cytochrome may correspond to a high molecular weight peroxidase with a protohaem IX. Such an enzyme has been isolated from *Bacillus stearothermophilus* (Loprasert *et al.*, 1988). A retention of protohaem, IX in cytochrome b subunit of the quinol oxidase from *Bacillus* PS3, has been observed in SDS gels, despite the boiling of the sample (Kutoh and Sone, 1988).

The other cytochrome, band three, has an apparent molecular weight of 34 kD. It may correspond to a common component, COII, of the caa3 and caa

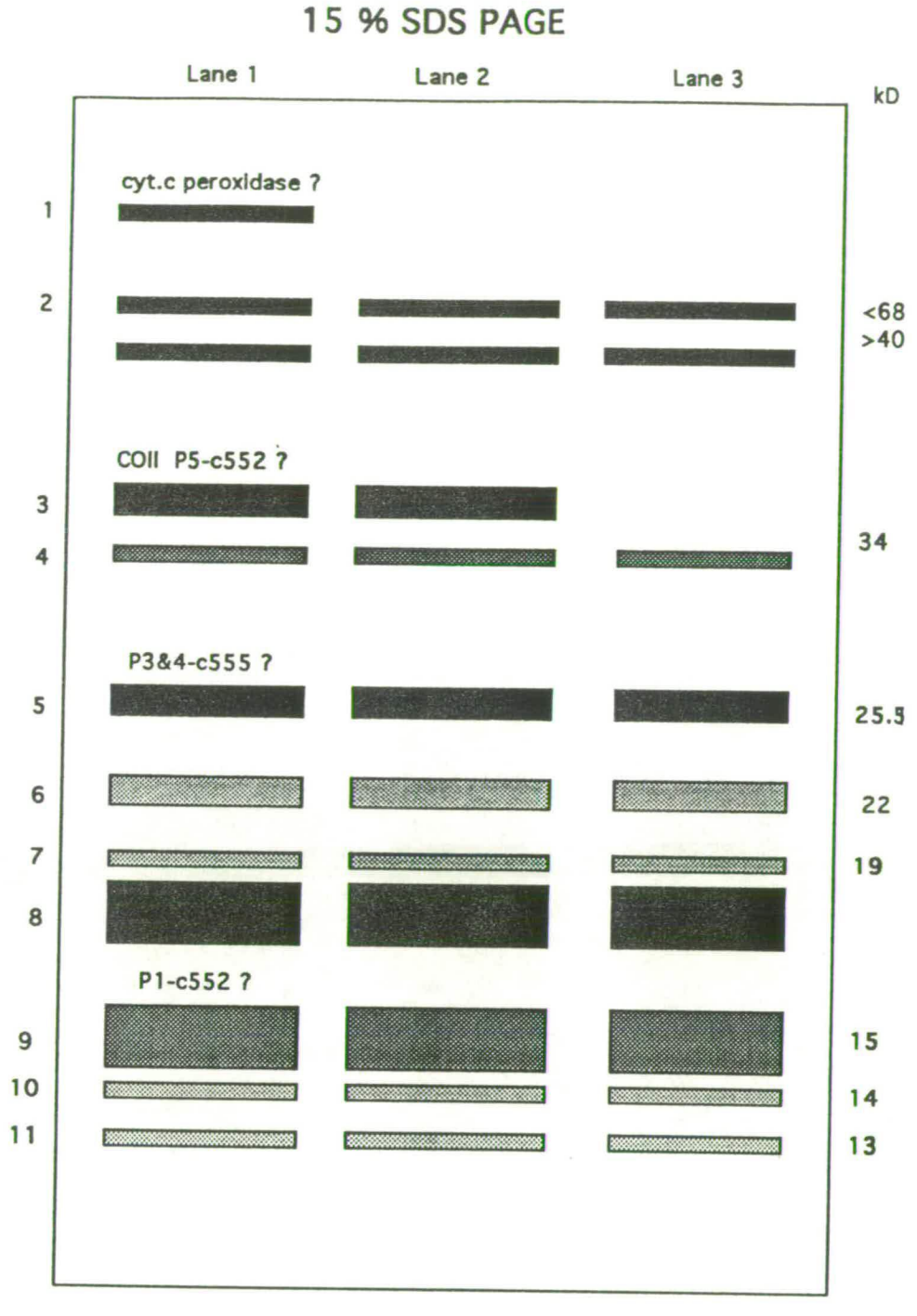


Figure 6.2. A schematic representation of the SDS 15% PAGE profile of the cytochromes from ***Bacillus azotoformans***, grown under different conditions.

Lane 1: Vigorous aeration.
Lane 2: Low aeration.
Lane 3: Anaerobic denitrifying conditions.

oxidases, as this band was observed both under well aerated and air-limited growth conditions. The oxidase in *Bacillus* PS3 has four subunits, COI (68 kD), COII, (38 kD), COIII (23 kD) and COIV (12 kD). The COII subunit has a covalently bound c-type cytochrome (Ishizuka *et al.*, 1990). It is also a component of the cao oxidase, which is derived from the caa3 oxidase under slightly air-limited conditions (Sone and Fujiwara, 1991). The aco oxidase is repressed under severely air-limited conditions in *Bacillus* PS3 and under anaerobic denitrifying conditions in various mesophilic *Bacillus* species (Pichinoty, 1978, Stouthamer, 1992). Accordingly, the COII subunit may be expected to be seen in samples from both well aerated cultures and cultures grown under slightly air limited conditions, but not in anaerobic denitrifying cultures. The expression of the 34 kD (band 3) cytochrome from *B. azotoformans*, conforms to this pattern.

One cytochrome c, P5-c552, which was isolated in this study and partially sequenced from the N-terminus, bears a slight resemblance to the caa3-cytochrome c from *Bacillus* PS3. They may be homologous. The P5-c552 cytochrome in *B. azotoformans* was, however, expressed under both aerobic and denitrifying conditions. It was obtained in low yields, but it may have been extensively degraded by the trypsin during the isolation procedure (see figure 4.10a,lanes 2 and 4). If this cytochrome is fused with a subunit of the caa3 complex in *B. azotoformans,* the fact that it is expressed under both denitrifying and aerobic conditions, may then best be explained on the basis of the genetic organisation of the operon coding for the caa3 cytochrome c oxidase genes in *Bacillus* PS3 (Ishizuka *et al.*, 1990). The gene sequence indicates that cytochrome c gene is a separate entity and even though the cytochrome c gene is fused to the C-terminal gene sequence of the COII subunit, it has the characteristic methionine start codon and a Shine Delgarno box homologue. There is also a stop codon between the cytochrome c and the oxidase subunit gene. However, while the chromosomal region between the stop codon and the start codon of the cytochrome c gene is G rich (a typical feature of a ribosome binding site), no palindrome region (terminator stem) can be found. Stop codon regions are typically rich in T and palindrome structures to facilitate the release of the ribosome. The observed fusion of the cytochrome c with the COII subunit is therefore either a posttranscriptional modification or, more likely, the translation of the COII subunit proceeds through the stop codon. If the P2-c552 cytochrome c of *B. azotoformans* is homologous to the *Bacillus* PS3 cytochrome, and if its operon has a similar organisation, the gene coding for the c-type cytochrome may then be expressed

independently and in the absence of any expression of other components of the caa3 oxidase complex. This cytochrome may also be a different structural variant of the same subclass as the caa3-cytochrome c from *Bacillus* PS3 and it may have a different role and be constitutive.

Electrophoretic studies and spectrophotometric analyses of cytochromes from *B. azotoformans*, did not reveal any specific induction of c-type cytochromes during denitrifying growth. The cytochromes, P1-c552, P2-c551, P3&P4-c555 split alpha and P5-c552, isolated in this study, were obtained in similar yields from both aerobic and denitrifying cultures of *B. azotoformans.*

In contrast to *B. licheniformis*, *B. azotoformans.* is a vigorous denitrifier. It has also five fold the total cytochrome c content of *B. licheniformis.* This abundance of cytochromes c may reflect this denitrifying activity ability and indicate a special adaptation to nitrate rich habitats. The corresponding genes for essential c-type cytochromes in this branch of the metabolic pathway, could also for the same reasons have become constitutive.

The cytochrome c obtained in the greatest yield, P1-c552 may correspond to band nine (approx. 15 kD), which appears to be completely digested. A cytochrome c of an apparent molecular weight of 15 kD has been sequenced from *Bacillus subtilis.* This protein apparently has two domains, one of which is a membrane spanning part.

A cytochrome with the same spectral properties as the P3&P4-c555 is apparently a part of the menaquinol reductase (bc554 complex) in *Bacillus* PS3. This cytochrome c has an molecular weight of 29 kD. The c555 split alpha cytochrome c from *B. azotoformans* may therefore correspond to the intense 25.5 kD band present in membrane samples from cultures grown both under aerobic and anaerobic denitrifying conditions.

The P2-c551 cytochrome was obtained in very small yields. It may be one of the three small haem staining bands, which were only partly degraded by trypsin (bands 8, 10 and 11). Cytochromes of this size are most likely to be cytochromes c. If two can be accounted for by the P2-c551 and the P5-c552 independently expressed from the caa3 oxidase operon, then it leaves the third one as not having been isolated. The faint bands not accounted for on the SDS-PAGE gels, of larger molecular weights, may be b-type or a-type cytochromes, which retain their haems to some extent..

The only induced cytochrome under anaerobic denitrifying conditions according to SDS-PAGE analyses, was a b-type cytochrome with unusual absorption spectra. Evidence for this is, however, tentative but this

cytochrome may be equivalent to the b-type cytochrome, which is induced under denitrifying conditions in *Bacillus halodenitrificans* (Denariaz *et al.*, 1991) and proposed to mediate electron transport from the menaquinones to the nitrate reductase.

6.2 PHYSICOCHEMICAL PROPERTIES OF CYTOCHROMES C FROM *BACILLUS AZOTOFORMANS*

Physicochemical properties commonly used for characterisation of cytochromes c include iron/haem content, ligand field of the iron, size of the protein, solubility, domain structure, spatial orientation, isoelectrical points, amino acid composition, absorption spectra/ extinction coefficient ratios, and redox potentials (Moore and Pettigrew, 1991). As these properties are more or less interrelated, readily obtained information on one particular property may give indirect information about another important property that is not as easily obtained directly.

By these means the broader classification of cytochromes c into classes is relatively facile but the further subdivision of the cytochromes into subclasses is a much more ambitious task.

6.2.1. Size and Amino Acid Composition

These 'properties', on their own, do not disclose much about the structural class of a corresponding cytochrome, but this is valuable supportive information. It can also contribute in classification of cytochromes of Class I along with other sequence characteristics, (Dickerson, 1980; table 6.2).

6.2.2. Isoelectrical Points

It has been proposed that isoelectric points correlated with Soret red./alpha red. ratios, with acidic proteins having high ratios and basic proteins having low ratios (Keilin, 1970). On the other hand they have been found nearly useless for classification purposes (Meyer, 1980).

Cytochromes c exhibit a wide variation in net charges within classes and within a particular subclass (Moore and Pettigrew, 1990).

Measuring the isoelectric points was deemed to be of no practical value for this work. The cytochromes c isolated from *B. azotoformans* appear to be only fragments of larger proteins and the isoelectric points cannot therefore be considered as true properties. The chromatographic behaviour and amino acid compositions indicate that these polypeptides are acidic. This was confirmed for cytochrome c552-P1 by sequencing. The cytochrome has a net negative charge of -6 (by discounting the fifth ligand histidine).

6.2.3. Haem Content

Haem content can be estimated directly by measuring either the iron or pyridine-haemochrome content in a cytochrome c of known molecular weight (Moore and Pettigrew, 1987). If the size and amino acid composition of the cytochrome under study are known, the haem content per cytochrome molecule can sometimes be inferred from the number of cysteines, histidines and methionines.

All the cytochromes c from *Bacillus azotoformans,* isolated in this study, have low molecular weight and each contains only a pair of cysteines. Thus it can be concluded that there is only one haem per molecule in all the cytochrome c polypeptides isolated and purified.

6.2.4. Ligand Field and Spectra

The exact determination of the ligand is ultimately dependent on X-ray crystallography. Indirect information can be obtained from amino acid composition, absorption spectra and magnetic characteristics by extrapolation from known aligned sequences and crystallographic structures (Moore and Pettigrew, 1990).

The fifth ligand is invariably histidine in the cytochromes c, characterised so far. The sixth ligand can be either histidine or methionine. Thus histidine is found as the sixth ligand in high-spin cytochromes c' of Class II whereas methionine is found in low-spin cytochromes c of Class I, II and III. Low-spin and high-spin cytochromes

are readily distinguished by their respective absorption spectra (Ambler, 1991).

The cytochromes P1-c552, P2-c551 and P5-c552 from *Bacillus azotoformans* have only one histidine. Cytochrome P3&P4-c555 has two histidines. All contain at least one methionine. Therefore, as all four cytochromes have characteristic low-spin spectra, it is probable that the fifth ligand is histidine and that the sixth ligand is methionine. This is supported by sequence information on P1-c552, P5-c552 and P3&P4-c555.

6.2.5. Spectra and Midpoint Redox Potentials

Spectral properties are sometimes used to classify cytochromes further into subclasses. Specific spectral properties are then known to correlate to a sequence subclass. For example, Bartsch (1991) has mapped, on the basis of spectral evidence as well as on available sequence evidence, the distribution of the various subclasses of cytochromes in a number of Gram negative bacteria (Bartsch, 1991). In Table 6.1 is compiled the available information on spectral properties and redox potentials of cytochromes c isolated from *Bacillus* species.

6.2.5.1. The cytochromes P1-c552 and P2-c551

Sequence information on *B. azotoformans* P1-c552 in this thesis (section 6.3.4) and on the *B. licheniformis* cytochromes c551 and c552 (Woolley, 1984, and van Beeumen, personal communication) reveals that they are homologous. They belong to the previously defined BacI subclass (section 6.1.2). The *Bacillus subtilis* cytochrome c550 and the *Bacillus* PS3 cytochrome c551, which have recently been sequenced (VonWachenfeldt and Hedterstedt, 1990; Fujiwara et al, 1993), also belong to this subclass, but their spectral properties have not been reported. Redox potentials have only been reported for the c552 and c551 sequences from *B. licheniformis*, the cytochrome P1-c552 *Bacillus azotoformans* (this work) and *B. halodenitrificans* cytochrome c550 (table 6.1). It is noteworthy that the redox potentials of the BacI cytochromes, except for the cytochrome c550 from *Bacillus* PS3, are all rather low, from to 100-153 mV. The redox potential of the *Bacillus* PS3 cytochrome c550 is reported to be 225 mV. This is surprisingly high, especially if one considers that the partial sequence of the *B. halodenitrificans* c550 cytochrome, which has 69% similarity to the corresponding region of the *Bacillus* PS3 cytochrome, has midpoint redox potential of only 130 mV. The sequence does not have any distinct

Table 6.1. Absorption maxima, extinction ratios and midpoint redox potentials of cytochromes c from *Bacillus* species.

	Type	Absorption maxima reduced spectra α nm	β nm	γ nm	280/ $alpha_{red}$	$Soret_{red}$/ $alpha_{red}$	$alpha_{red}$/ $beta_{red}$	Midpoint redox potential mV
B. azotoformans								
P1-c552	BacI	552	523	415	0.70	5.61	1.61	122
P2-c551	(BacI?)	551	522	416	0.84	6.21	1.69	93
P4-c555	Split alpha	555	523	417	0.72	6.56	1.38	124
P5-c552	(c2 like?)	552	524	417	0.78	5.14	1.66	188
B. licheniformis								
c552	BacI	552	522	416	1.13	6.25	1.39	124
c551	BacI	551	522	416	1.11	5.89	1.69	153
c554	Split alpha	554	522	417	1.13	6.94	1.10	136
B. subtilis								
c550	(BacI?)	550	520	414	1.20	5.00	1.95	220
c554	Split alpha	554	521	417	1.18	7.18	1.35	-83
Bacillus PS3								
c551	BacI	551	522	416	--	6.10	--	225
c553	Split alpha	553	--	--	--	--	--	200
B. halodenitrificans								
c550	BacI	550	--	--	--	--	--	130

Information on spectral properties and redox potentials are found in the following references: *Bacillus azotoformans* (this work; *Bacillus licheniformis*, (Woolley, 1986); *Bacillus subtilis*, (Miki and Okunuki, 1969), *Bacillus* PS3, (Sone et al., 1983, Soneet al., 1987), *Bacillus halodenitrificans*, (Sariva et al., 1992). Defination of type designations are found in text in sectons. -- means that no information was available.

differences from the other BacI sequences, which would help to explain this considerable higher redox potential. The *B. subtilis* c550 cytochrome (Miki and Okunuki, 1969) is reported to have a midpoint redox potential of 225 mV. This value is comparable to that of the *Bacillus* PS3 cytochrome c551, but as the relatively high alpha/beta extinction ratio, 1.95, also falls outside the range of values obtained for sequenced cytochromes of the BacI subclass, it may have a different structure from that of the BacI sequences. However, care should be taken in reading too much out of similarities or dissimilarities between spectra or midpoint redox potentials, especially when the data is limited.

The spectral properties of the P2-c551 cytochrome from *B. azotoformans* fall within the range of values observed for the BacI cytochrome. It may, therefore, also be of this class. This is supported by the measurement of the midpoint redox potential, 93 mV. Amino acid analysis revealed that tryptophan was present in the sequence that also supports the classification. Tryptophan is highly conserved near the C-terminus of the sequenced BacI cytochromes.

6.2.5.2. The cytochrome P5-c552 from *B. azotoformans*

The P5-c552 cytochrome has similar spectral properties to the sequenced BacI cytochromes c, but the midpoint redox potential is significantly higher (200 mM). P5-c552 may therefore belong to a different subclass, or be a different structural variety of, or related to the BacI cytochromes. The limited sequence information on this protein (section 6.3.2) indicates that this is a different cytochrome c type and perhaps more similar to the caa3-cytochrome c from *Bacillus* PS3.

6.2.5.3. The cytochrome P3&P4-c555 *B. azotoformans*

P3&P4-c555 has a characteristic split alpha peak. Cytochromes with similar spectral properties are found in the related bacteria, *Bacillus PS3*, *B. subtilis* and *B. licheniformis* (see section 6.1.3.1). Cytochromes, with similar spectra, are also found in purple bacteria. Sequence information has been obtained on such cytochromes c, from *Paracoccus halodenitrificans* and *Thiobacillus neopalitanus*, and the dihaem c4-cytochromes. Together they constitute the sequence subclass IC of Class I cytochromes c (Ambler, 1991).

The point worth noting, is that the midpoint redox potential of the split alpha cytochrome P3&P4-c555 is considerably higher than was reported for the split alpha cytochrome c554 from *B. subtilis* (Miki and Okunuki,

1969). It is 124 mV, compared with -80 mV for the *B. subtilis* cytochrome. The latter value is unusually low for a c-type cytochrome and lower values are only found among the c3 cytochromes from *Desulfovibrio* (approximately -200). A still higher midpoint redox potential, 200 mV, has been reported for the split alpha cytochrome c, in *Bacillus* PS3 (Sone *et al.*, 1983), which they propose forms a quinol reductase complex with b-type cytochrome c. The alpha peak of this cytochrome is reported to be at 553 nm.

The similarities of the absorption spectra of P3&P4-c555 to the IC subclass cytochromes from purple bacteria may be the result of convergence of tertiary nature rather than to a conservation of some common conserved ancestral character. Variations in spectra may primarily reflect haem attachment character, strain on the ligand field structure and the degree of shielding of the haem. The same or similar spectral versions may therefore be attained in a variety of ways by cytochromes c of widely divergent sequences.

The sequencing of the P3&4-c555 provides an opportunity to see if similar spectral properties between distantly related cytochromes are coded in the same way into the primary structure or, in other words, do these special spectral properties correlate to recognisable sequence motifs. An example of this, are the euglenoid cytochromes c558, where the considerable red shift of the alpha peak (558 nm), the result of only one thioether link, is evident in the sequence. The haem site has, correspondingly, only one cysteine. This may be an unique example or rather, that such a clear correspondence of a spectral character and a sequence motif is seldom to be found.

6.2.6. Solubility and Domain Structure of Cytochromes c from The Genus *Bacillus*

It is now generally accepted that cytochromes c are located on the outer side of the cytoplasmic membrane. In Gram negative bacteria they are periplasmically retained and can usually be obtained as readily soluble entities. The Gram positive bacteria lack a periplasmic space and the cytochromes c must therefore require specific adaptations for attachment to the membrane, so as not to diffuse into the surroundings. Because of small quantities of cytochromes c in many Gram positive bacteria and the difficulties encountered in solubilising them, they have received scant

attention. Their mode of attachment is still unclear and the comparatively few reports hitherto published, differ in their conclusions.

In the present study, attempts were made to clone and sequence genes, which code for cytochromes c from *Bacillus licheniformis* with the aim, amongst others, to find out how they might be attached to the membrane. Unfortunately, this was unsuccessful. However, information obtained on the four cytochromes c isolated in this work from *Bacillus azotoformans* is relevant to this discussion. SDS electrophoresis of the membrane fraction not treated with protease, shows that no native cytochrome c is smaller than 12 kD. The four isolated cytochromes were obtained by proteolysis and their approximate sizes, as determined by SDS electrophoresis, are from 8 to 10 kD. The P1-c552, P3&P4-c555 and the P5-c552 cytochrome apparently have an extended N-terminus, which does not form a part of the cytochrome core, but still, according to electrophoresis analyses, even the largest polypeptide must have lost least 20 amino acids. All of the isolated cytochromes are big enough to be complete functional haem domains. This is indeed supported by sequence information on at least two of them, P1-c552 and P3&P4-c555. The isolated cytochromes c from *B. azotoformans* have apparently been cleaved off their N-terminal membrane anchorage parts and are most likely integral membrane proteins or covalently fused to other proteins.

In this laboratory soluble cytochromes had been obtained from *B. licheniformis* by proteolysis (Thatcher and Ambler, unpublished observation,), which by itself indicated that they had been severed from a membrane anchorage moiety. This was followed up by Woolley (1984, 1987). Although he only obtained soluble cytochromes by mild proteolysis, he concluded that they were not membrane bound as he only observed single bands on SDS electrophoresis gels.

The two cytochromes c which Miki and Okunuki (1969) obtained from a *Bacillus subtilis* strain, were in the cytoplasmic fraction. They concluded that these cytochromes c were soluble entities.

The gene sequences and analyses of the mature forms of the BacI cytochromes c in *Bacillus subtilis* (VonWachenfeldt and Hederstedt, 1990) and *Bacillus* PS3 (Fujiwara et., 1993), indicate that their mode of attachment of to the membrane is different. The gene for the cytochrome c550 from *Bacillus subtilis* appears to code for an 120 amino acid long, two domain cytochrome. The calculated weight is 13 kD, which corresponds to the measured weight on SDS electrophoresis gels. The smaller domain is 21 amino acids long, hydrophobic and appears to be an alpha helical

segment, spanning the membrane. In contrast the cytochrome c551 from *Bacillus* PS3 appears to be translated into an 110 amino acid long polypeptide, which is then processed into a shorter form of 93 amino acids. Its apparent molecular weight on SDS-PAGE is 10.5 kD. The protein is blocked at the N-terminus. The authors note that the retention time of the protein by reverse phase chromatography, before and after treatment with lipase, is different, and they propose that the cytochrome is acylated at the N-terminus. In support of this, it can be pointed out that there is a tetrapetide sequence -LAAC- in the signal peptide. A similar tetrapeptide, -LAGC-, is a post-translational modification site in the signal peptide of membrane bound penicillinases in *B. licheniformis* and in *E. coli* (Lai *et al.*, 1981, Nielsen *et al.* 1981). A cystyl diacyl glycerol is found at the N-terminal and the amino group of the leucine residue at the end is also acylated. The preceding signal sequence is cleaved off. The partially sequenced cytochrome c550 from *Bacillus halodenitrificans* has a very similar N-terminal sequence, but it has a shorter N-terminus starting with the two glycines, which appear to be conserved between them. The apparent molecular weight of the *Bacillus halodenitrificans* cytochrome c550 on SDS-PAGE is 9.5 kD. It is not reported to be acylated. The cytochrome might be soluble, but a concomitantly expressed protease might also have severed the haem domain from its membrane attachment part during the isolation procedure. These two *Bacillus* cytochromes are also more similar, in the haem domain sequence, to each other (69% identity) than each is to any other cytochrome of the BacI subclass.

A cytochrome, structurally different from the BacI sequences, was found, by DNA sequencing, to be a part of the caa3 oxidase operon in *Bacillus* PS3. Electrophoretic analysis on SDS-PAGE indicates that it is fused to one of the subunits (COII).

From the above, it appears that cytochromes c may be retained on the external surface of the *Bacillus* cell membrane, in a number of ways. First, it may be bound, covalently or otherwise, in a functional complex with an integral membrane protein, such as may be the case in the caa3 oxidase complex in *Bacillus* PS3 (Ishizuka *et al.*, 1990). Second, it can be an integral membrane protein with two domains one of which is a hydrophobic membrane spanning segment, as in *Bacillus subtilis* cytochrome c550 (VonWachenfeldt *et al.*, 1991). Third, it may be a peripheral membrane protein with postranslational modifications, such as an N-terminal acyl group as is proposed for the PS3 cytochrome c551. These modifications may help to anchor the protein to the membrane.

It can also be mentioned that proteins on the outer surface of the membrane encounter the cell wall before they are free to diffuse away. The Gram positive cell wall is a highly cross linked semi-porous structure, composed primarily of peptidoglycan and anionic polymers. Consequently, the passage of proteins through it may be restricted by antagonistic charge interactions under the prevailing physiological conditions.

I can further add that in recent work, while isolating unrelated enzymes from thermophilic *Bacillus* species, I noticed characteristic cytochromes c absorption peaks in the soluble fraction. Whether these are proteolytic fragments or truly soluble cytochromes, is not known. The quantity was not great considering that the bacteria were grown in 100 litre batches.

6.2.7. Conclusions

While physicochemical properties are useful for establishing the identity of a particular cytochrome c and for functional analysis, they are of limited value for classification purposes. This is best achieved on the basis of primary and tertiary structures. When such information is not available, valid proposals can often be made by extrapolation from known primary structures and what is known about their physicochemical properties, especially from their absorption spectra and midpoint redox potentials. Many physicochemical properties can be read more or less directly from primary and tertiary structures. Others, such as spectral characteristics and redox potentials, are certainly coded in those structures, but their code has not been broken convincingly yet, but for the most generalised principles. It can be argued, that it has a little intrinsic value and that it is impossible for minor spectral variations, but if the properties bear relation to the functional attributes of the protein, it is of an obvious merit, as this information is immediately accessible.

Physicochemical analysis of cytochromes c isolated from *B. azotoformans* reveals that all four probably belong to Class I, and that they appear to be complete haem domains of larger integral membrane structures. The haem polypeptides are in the lower size range of the soluble cytochromes encountered in Gram negative bacteria. Spectral properties may also have been defined for the BacI subclass cytochromes c from Gram positive bacteria (table 6.3).

The following conclusion can also be drawn or reinforced from the results obtained in this work: Spectral properties, defined for a sequence subclass from a particular group of bacteria, cannot with any certainty, be predictive of sequence class from phylogenetic distant groups.In this study it was observed: Firstly, the spectral properties and midpoint redox potentials of the ID subclass like cytochromes of Gram positive bacteria, while similar, are different from those of their counterparts in purple bacteria, despite the considerable sequence homology. Secondly: Cytochromes c more divergent in primary structure, and possibly more putitativily of the same subclass, the P3&P4-c555 cytochrome and the IC subclass cytochromes, have distinctly similar spectras.

6.3. CYTOCHROMES C SEQUENCES FROM *BACILLUS AZOTOFORMANS*

Physicochemical data for all four cytochromes c isolated from *Bacillus azotoformans* and sequence information on three of them indicate that these cytochromes all belong to the Class I (sections 6.5.2 & 6.5.3. Table 6.2).

6.3.1. Cytochrome P2-c551

The least information was obtained on the P2-c551 cytochrome c, only of the N-terminus. The sequenced part does not have any characteristics that might help to classify it, but it is probably part of an N-terminal linker sequence that connects the haem domain to its membrane anchorage part (see section 5.3). The physicochemical properties indicate that this cytochrome c sequence may be of the BacI subclass.

6.3.2. Cytochrome P5-c552

A 37 amino acids long N-terminal sequence was obtained from the P5-c552 cytochrome. Figure 6.3 shows the P5-c552 sequence aligned with the caa3-cytochrome c from *Bacillus* PS3, the caa3-cytochrome c from *Thermus thermophilus*, the P1-c552 from *Bacillus azotoformans* and the c2 cytochromes from Tuna, *Rhodospirillum acidophila* and *Rhodopseudomonas spheroides.* The sequence is aligned with representatives of all subclasses in figure 6.17a. It shows resemblance to

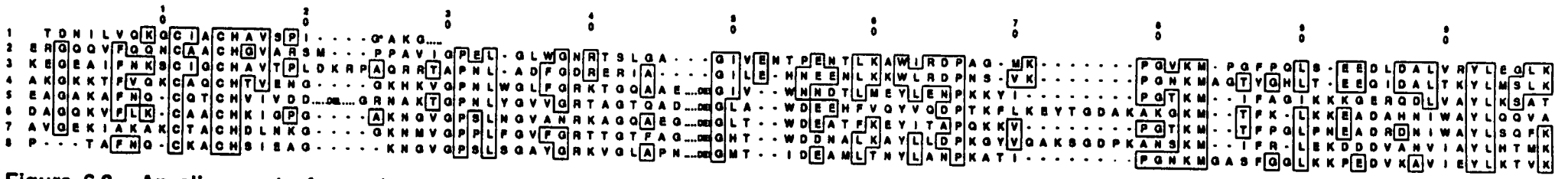


Figure 6.3. An alignment of cytochrome P5-c552 from *Bacillus azotoformans* and *Bacillus* PS3 to selected c2 cytochromes.

G* in position 25 could also be arginine. Del : the corresponding sequence part is not shown. Amino acids identical to those in the *Bacillus* PS3 c(aa3) sequence (nr. 3) are boxed.

1: Partial sequence of cytochrome P5-c552 from *Bacillus azotoformans* (this work).
2: Cytochrome caa3 from *Thermus thermophilus* (Mather, 1991).
3: Cytochrome caa3 from *Bacillus* PS3 (Ishizuka et al., 1990).
4: Tuna cytochrome c (Kreil et al., 1963, 1965 (cited in Moore and Pettigrew, 1990)).
5: Cytochrome c2 from *Rhodobacter sphaeroides* (Ambler, 1979).
6: Cytochrome c2 from *Rhodopseudomonas acidophila* (Ambler, 1979).
7: Cytochrome c2 from *Rhodospirillum photometricum* (Ambler, 1979).
8: Cytochrome c2 from *Rhodospirillum fulvum* iso-1 (Ambler, 1979).

the caa3-cytochromes c from *Bacillus* PS3 and *Thermus thermophilus* in the region around the haem site, but it also has substantial differences. A glycine rich region immediately follows the haem site region, which may indicate a turn or a loop. This feature is not found in the caa3 cytochrome c sequence from *Bacillus* PS3. Also, the long N-terminal sequence preceding the haem site does not bear any resemblance to the corresponding part of the caa3 cytochrome c from *Bacillus* PS3 (see sections 5.4 and figure 6.6). It is noteworthy that the *Bacillus* caa3 sequence is obviously homologous to the *Thermus thermophilus* caa3 sequence, with 40% identity (figure 6.17b) and these two sequences show the greatest similarity to the c2 cytochrome sequences of purple bacteria of all Class I sequences (sections 6.3.3 & 6.5.4.2, figures 6.3 & 6.17a).

6.3.3. Cytochrome P3&P4-c555

Cytochrome P3&P4-c555 was sequenced to a large extent. Figure 6.4 shows an alignment of the sequenced peptides to the c4 cytochromes and a selection of algal cytochromes c (soluble f cytochromes). The sequence parts obtained cover the N-terminus region including the haem site, the proline site and also the C-terminus region, including the sixth ligand methionine. Another peptide sequence which sequenced cleaner than the first one, 24 amino acids, was also obtained, but it did not have overlap with any of the terminal peptides. Its position in the protein is therefore uncertain. Its proposed but highly conjectural place in the structure is shown in figure 6.4.

Even by taking into account only the sequenced N-terminus and C-terminus regions, it can be stated that the primary structure of the P3&P4-c555 cytochrome is different from any of those which have been sequenced previously from the genus *Bacillus*. By aligning the sequence with cytochrome sequences from Gram negative bacteria, an affinity was found to the c4 cytochromes of the IC subclass. The c4 cytochromes group, consists of the split alpha monohaem cytochromes of *Thiobacillus neapolitanus* and *Paracoccus denitrificans*, and both the domains of the dihaem cytochromes c4 of *Azotobacter vinelandii* and *Pseudomonas aeruginosa*. Interestingly, the P3&P4-c555 cytochrome also has a split alpha peak.

The *B. azotoformans* partial sequence shares approximately 25% identity with comparable regions of each of the c4 sequences. The similarity is

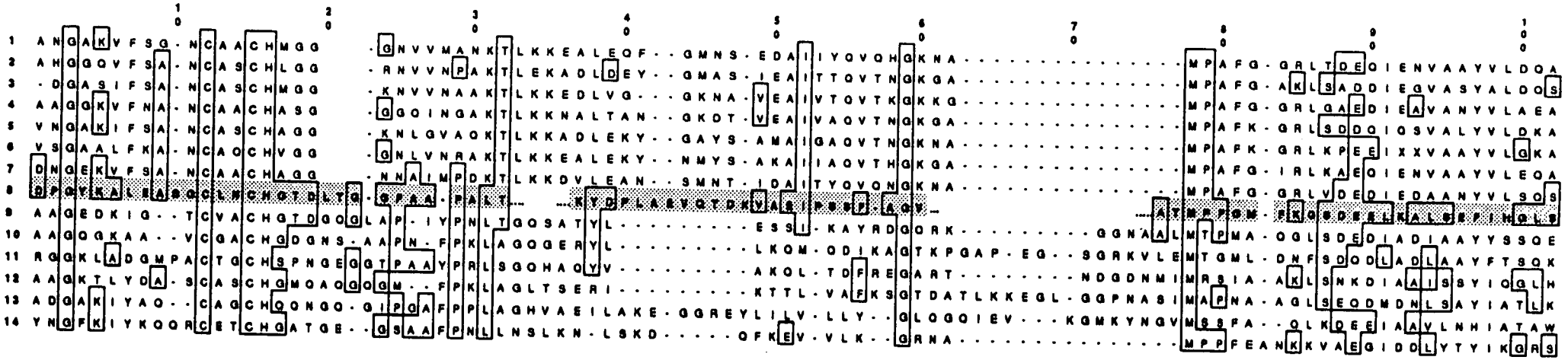


Figure 6.4. **Alignment of the partial sequence of P3&4 c555 from *Bacillus azotoformans* to the algal (soluble f) cytochromes, the c4 cytochromes cytochrome c552 from *Thermus thermophilus* and the cytochrome c555 from *Methylococcus capsulata.***

The partial sequence of the P3&4 cytochrome c from *Bacillus azotoformans* is in boldface.
Amino acids, which are identical to those in the *B. azotoformans* sequence, are boxed.

1 ***Synechococcus*** sp. (Aitken, 1976).
2 ***Anacystis nidulans*** (Borden and Margoliash, cited in Moore and Pettigrew, 1990).
3 ***Microcystis aeruginosa*** (Beecher and Margoliash, citedin Moore and Pettigrew, 1990).
4 ***Plectonema boryanum*** (Aitken, 1976).
5 ***Anabaena variabilis*** (Beecher and Margoliash, cited in Moore and Pettigrew, 19!
6 ***Aphanizomenon flos-aquae*** (Sprinkle et al ., cited in Moore and Pettigrew, 1990).
7 ***Porphyra tenera*** (Ambler and Bartsch, 1975).
8 ***Bacillus azotoforman:*** P3&4-c555 (Hreggvidsson, this work).
9 ***Paracoccus denitrifican*** (Ambler and Bartsch, 1975).
10 ***Azotobacter vinelandii*** (Ambler and Bartsch, 1975).
11 ***Azotobacter vinelandii*** (Ambler and Bartsch, 1975).
12 ***Thibacillus neapolitans*** (Ambler and Bartsch, 1975).
13 ***Thermus thermophilus*** (Titani et al. 1985)
14 ***Methylococcus capsulata*** (Ambler et al. 1986).

strongest in the region following the haem site. Notable is the retention of highly conserved pair of prolines, one of which probably hydrogen bonds to the fifth ligand histidine. These prolines are one of the most prominent features of the IC subclass as noted by Ambler (1977, 1982, table 6.2). In the C-terminus of the P3&P4-c555 cytochrome there are identical or structurally similar residues in many of the homologous and conserved positions of the c4 sequences. There are also some notable differences. The major ones are in positions 81, 84 and 86 (figure 6.4), where the residues do not follow the observed hydrophobic pattern of the c4 sequences. The presence of glycine in position 86 is most discordant. In this position a hydrophobic amino acid is found in all other cytochrome sequences. Also, there is a phenylalanine in position 97, where there is a predominant conservation of tyrosine in the c4 group. It can perhaps be stated that, in the C-terminal region, the split alpha cytochrome c from *B. azotoformans* has a distinct character of its own with no greater affinities to any one cytochrome c type rather than another.

The IC subclass as defined by Ambler (1977, 1982) also includes the algal cytochromes (c6) but the similarity to other cytochromes of the IC subclass is almost exclusively in the C-terminal region. This is evident from the alignment in figure 6.4 (see also section 6.5.4.2).

Sone *et al.* (1987) have observed that the c554 split alpha cytochrome in *Bacillus* PS3 forms a complex with a b-type cytochrome, and that it functions as a menaquinol-cytochrome c reductase. They have proposed that the split alpha cytochrome c is a c1 or f-type cytochrome. These particular cytochromes are found in the b6f and bc1 complexes of cyanobacteria and chloroplasts and, purple bacteria and mitochondria, respectively. The menaquinol cytochrome c reductase from *Bacillus* PS3 has a comparable physiological role and properties, but it is more similar to the b6f than to the bc1 complex.

The reported c1 sequences from mitochondria and purple sulphur bacteria and the f cytochrome sequences from chloroplasts have a very distinct character and cannot, with any confidence, be placed with cytochromes c of Class I. They may form a class of their own. On the other hand, the partial sequence of the split alpha P3&P4-c555 includes sequence parts which can be aligned with considerable certainty to N- and C-terminal sequence parts of proper Class I cytochrome c sequences.

The only cytochrome f sequence from cyanobacteria reported to date is from *Nostoc* sp. It is very similar to the spinach cytochrome c1 sequence with amino acid identity of 69% (Kallas *et al.*, 1988). It must be pointed out

that the cytochromes f from chloroplasts and cyanobacteria are different from the soluble f cytochromes in the same organisms. The latter ones, also called algal cytochromes, are small cytochromes that belong to Class I.

The P3&P4-c555 split alpha cytochrome can perhaps be grouped with the IC subclass sequences as they are defined by Ambler (1982,1991). It has the greatest similaritiy with the c4 cytochromes, but only slight similarity with the algal soluble f cytochromes. More accurate classification must await the full sequence.

6.3.4. Cytochrome P1-c552

A complete sequence was obtained of the haem domain of the P1-c552 cytochrome. Figure 6.5a shows the P1-c552 cytochrome aligned with homologous cytochromes from Gram positive bacteria and the ID subclass sequences from purple bacteria and the algal cytochromes (soluble f). The distance matrix for this alignment is shown in figure 6.5b. Figure 6.17a shows the alignment of the P1-c552 cytochrome to a selected set of cytochromes c, which includes representatives of all the defined subclasses of Class I and a few unclassified sequences as well.

The cytochromes c from the Gram positive bacteria are clearly of the same structural type. They have a minimal identity of approximately 40%. There is also a marked similarity between these cytochromes and the ID subclass cytochromes. 75% of the sequence identity scores to 14 ID subclass sequences are equal or higher than 30%. In contrast, only 19% of identity scores with the 8 most similar algal (soluble f) sequences have values equal or higher than 30%.

Particularly noteworthy is the conserved pattern of amino acids, methionine (position 69), proline (position 73) and valine (position 79) (figure 6.5a) and the conserved tryptophan (position 90) close to the C-terminus between the BacI and the ID subclasses. The hydrophobic periodicity of the helix regions of ID subclass also appears to be conserved.

Fujiwara *et al.* (1993) have emphasised a special relationship of the PS3-c551 cytochrome sequence with that of c6 algal cytochromes. In my opinion this proposal is unwarranted, for the reasons given above, and because too many indel events (figure 6.6.) have been proposed in their alignment compared with the alignment to the most similar ID subclass sequences. Furthermore, by pairwise profile consensus scoring (section 6.5.4.3, appendix B, figure 6.21), affinity of 0.64 (lowest score=0, highest

score=1.00) is between the ID and the BacI subclasses while only 0.46 was found between the algal and the BacI subclasses. The latter is comparable to such as those between the algal and c4 sequence groups (0.46) and the BacI and the c2 sequence groups (0.48).

The BacI sequences have distinct characteristics of their own, which set them apart from the ID subclass cytochrome sequences from the purple bacteria. The most notable difference between these two sequence groups, are the two deletions, four amino acids from positions 39 to 42 and 6 amino acids in the region from position 58 to 68, in the BacI sequences (Figure 6.5a), corresponding to two small glycine-rich regions, one of which includes a very well conserved tryptophan, in the purple bacteria sequences. Perhaps just as noteworthy, are the differences in the loop structure sequence following the haem site and the presence of glycine in the second helix region (in position 33, Figure 6.5a),both of which indicate that these regions have different conformations in the two subclasses. These characteristics do not point at any special relationship with the algal cytochromes, either. A heptapeptide sequence adjacent to the haem site, is more similar sequentially to a corresponding region in the P3&P4-c555 sequence. This may represent a convergent evolution or a shared ancient feature, but another explanation might be a recombination event between ancestral sequences of these *Bacillus* cytochromes. The sequences are CHGQNLEGG in the P1-c552 cytochrome and CHGTDLTGG in P3&P4-c555 cytochrome.

The *Bacillus* sequences certainly bear a strong resemblance to the ID subclass cytochromes c from purple bacteria, but whether to form a subclass with the ID sequences or create a specific subclass, is a matter of definition. In this thesis the structural type is designated BacI or subclass BacI.

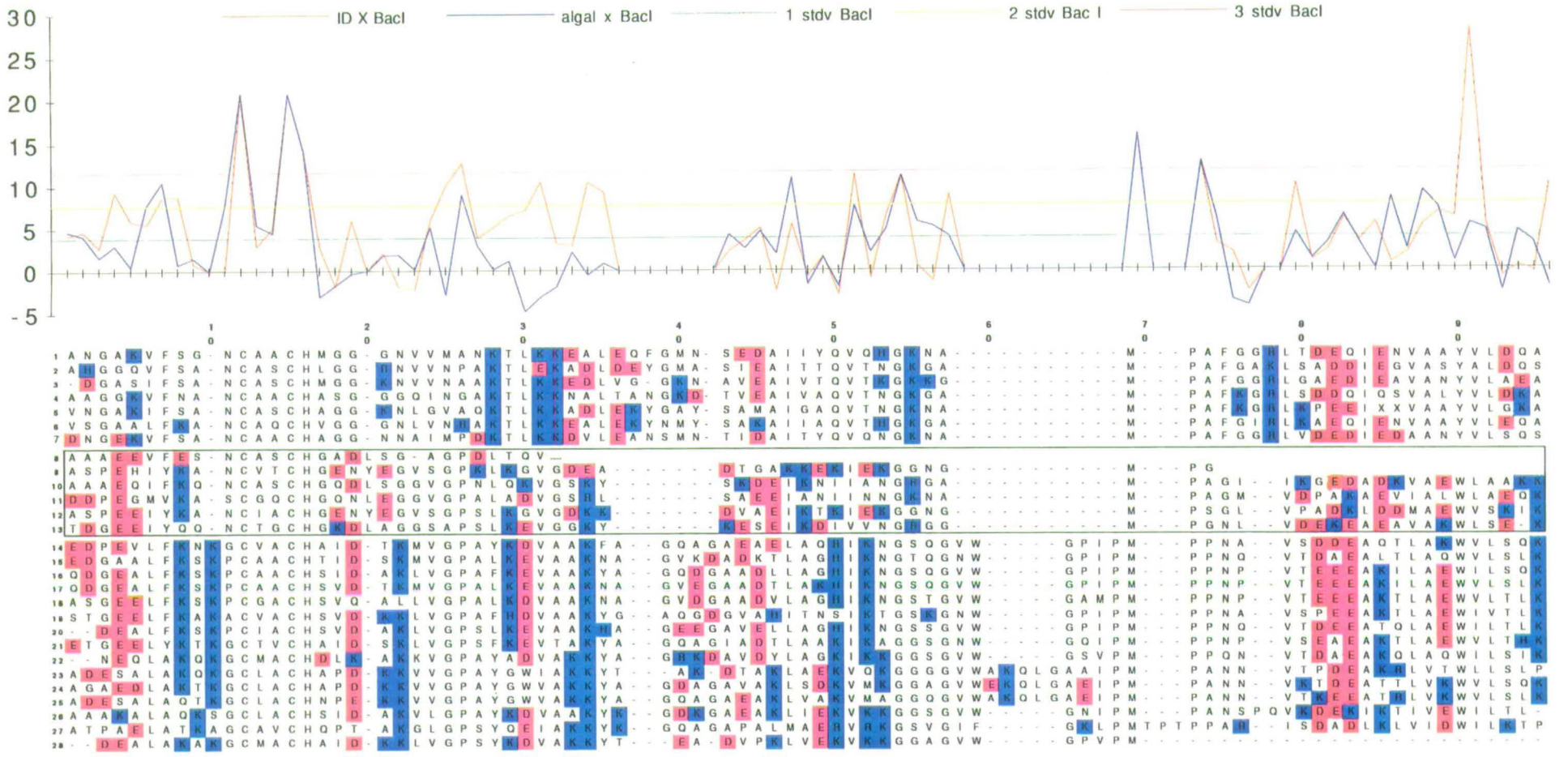


Figure 6.5a. Alignment of cytochromes c of the BacI subclass *Bacillus*, the algal (soluble f) cytochromes and the ID subclass cytochromes.

The BacI sequences are boxed in. Graphical representation of the positional profile scores (Dayhoff 'units' 100 PAM) between the consensus sequence of the BacI and the consensus sequences of the ID- and the algal subclasses is presented above the alignment. The plots reflect the positional conservation between the BacI subclass and the algal- and the ID subclasses, respectively (see section and 6.4.1 and appendix B).

Negatively charged amino acids
Positively charged amino acids

1 *Synechococcus* sp (Aitken, 1976).
2 *Anacystis nidulans* (Borden and Margoliash, cited in Dickerson, 1980).
3 *Microcystis aeruginosa* (Beecher and Margoliash, cited in Ulrich et al., 1982).
4 *Plectonema boryanum* (Aitken, 1976).
5 *Anabaena variablis* (Beecher and Margoliash, cited in Ulrich et al., 1982).
6 *Aphanizomenon flos-aquae* (Sprinkle et al ., cited in Ulrich et al., 1982).
7 *Porphyra tenera* (Ambler and Bartsch, 1975).
8 *Bacillus halodenitrificans* (Sairiva et al., 1992).
9 *Bacillus licheniformis* (Wolley (1984) and VanBeeuman personal communication).
10 *Bacillus* PS3 (Fujiwara et al., 1993).
11 *Bacillus azotoformans* (Hreggvidsson this work).
12 *Bacillus subtilis* (VonWachenfeldt et al., 1990).
13 *Bacillus licheniformis* (Wolley and VanBeeuman personal commun.).
14 *Pseudomonas aeruginosa* (Ambler, 1963).
15 *Pseudomonas fluoroscens* (Ambler, 1991).
16 *Pseudomonas stutseri 221* (Ambler, 1991).
17 *Pseudomonas stutseri 320* (Ambler, 1991).
18 *Pseudomonas mendocina* (Ambler, 1991.)
19 *Pseudomonas denitrificans 9496* (Ambler, 1991).
20 *Pseudomonas denitrificans 10465* (Ambler, 1991)
21 *Azotobacter vinelandii* (Ambler, 1991).
22 *Hydrogenobacter thermophilus* (Sanbogi et al., 1989).
23 *Rhodocyclus purpureus* (Ambler 1991).
24 *Rhodospoirillum tenue 2781* (Ambler, 1991).
25 *Rhodospirillum tenue 3781* (Ambler, 1991).
26 *Methylophilus methylotrophus* (Ambler, 1991).
27 *Rhodopseudomonas gelatinosa* (Ambler, 1991).
28 *Pseudomonas hydrogenothermophilus* (Kodama, 1992).

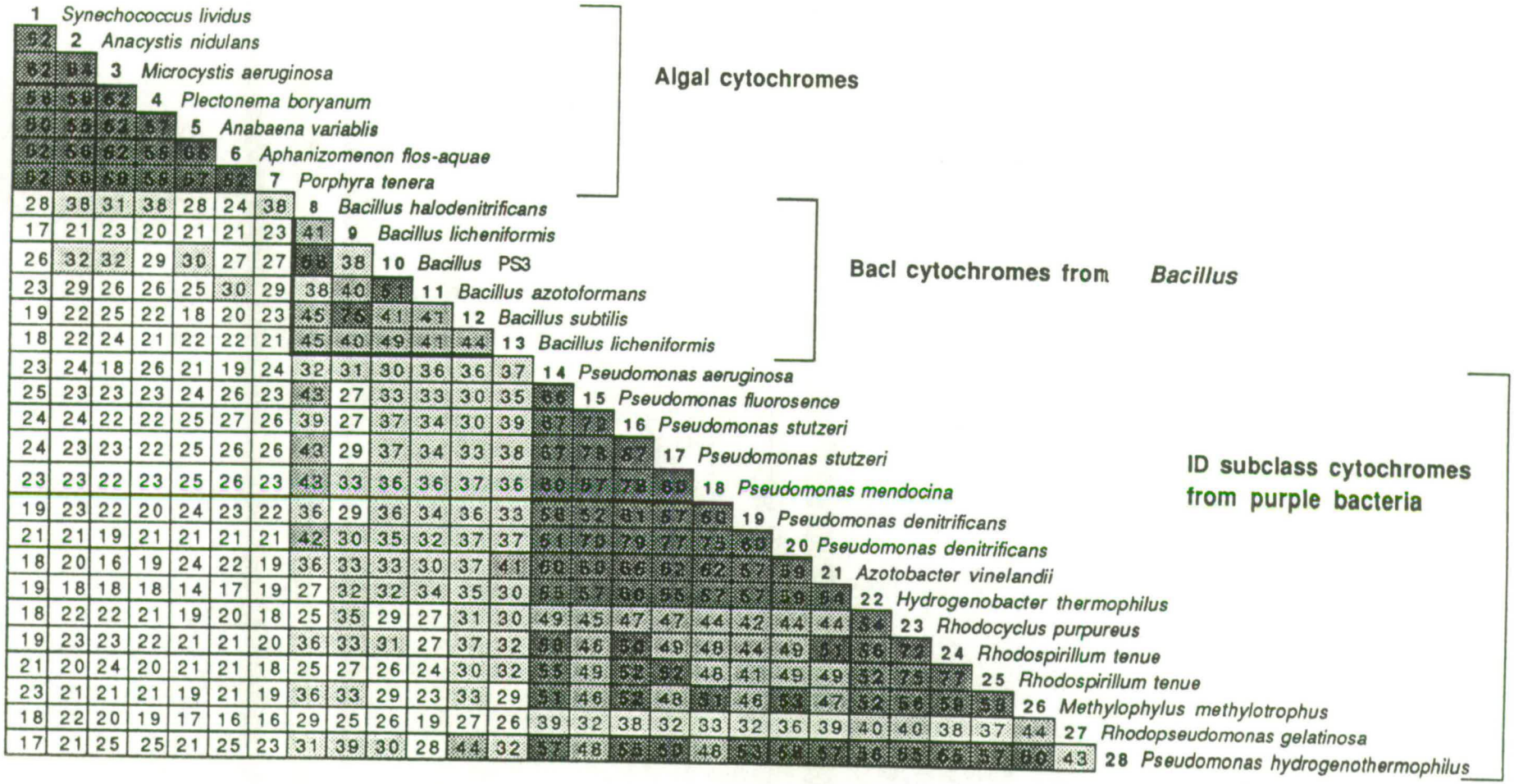


Figure 6.5b. Matrix of % identities for the alignment in figure 6.5a of cytochromes c of the BacI subclass from the genus *Bacillus*, the ID subclass from the purple bacteria and the algal (soluble f) cytochrome c subclass.

The values are percentage identities of minimum common sequence length.

similarity values 50% and higher. similarity values 40% to 50%. similarity values 30% to 40%.

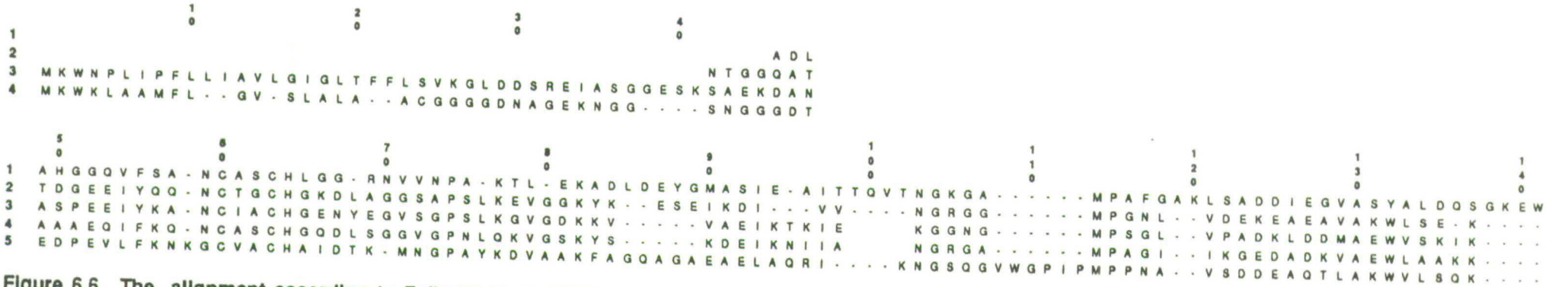


Figure 6.6. The alignment according to Fujiwara et al. (1993) of Bacl subclass cytochromes c to ID and algal subclass cytochromes c.

1 *Anacystis nidulans* (Borden and Margoliash, cited in Dickerson, 1980).
2 *Bacillus licheniformis* (Wolley and VanBeeuman personal communication).
3 *Bacillus subtilis* (VanWachenfeldt et al., 1990).
4 *Bacillus* PS3 (Fujiwara et al., 1993).
5 *Pseudomonas aeruginosa* (Ambler, 1963).

6.4. STRUCTURAL ANALYSIS OF THE P1-C552 CYTOCHROME C FROM *BACILLUS AZOTOFORMANS*

6.4.1. Profile Analysis

Positional identity of amino acids is too severe a condition for asserting structural relationship between proteins. Amino acids can substitute for each other and this depends on their physicochemical character and the specific position of the amino acid in the protein. In figure 6.7. is a Venn diagram made by R.W. Taylor (1986), which shows the relationships of the 20 amino acids to a selection of physicochemical properties, size, charge, aliphaticity, hydrophobicity and polarity, that are important in the determination of protein tertiary structure. In other words this Venn diagram shows the degeneracy of the code relating the amino acids to the tertiary structure and function of proteins. These relationships between amino acids are also reflected in the relatedness odds matrix or similarity matrix of Dayhoff et al., (1978). This matrix is based on the observed rate of mutations within a family of closely related proteins.

Figure 6.7.

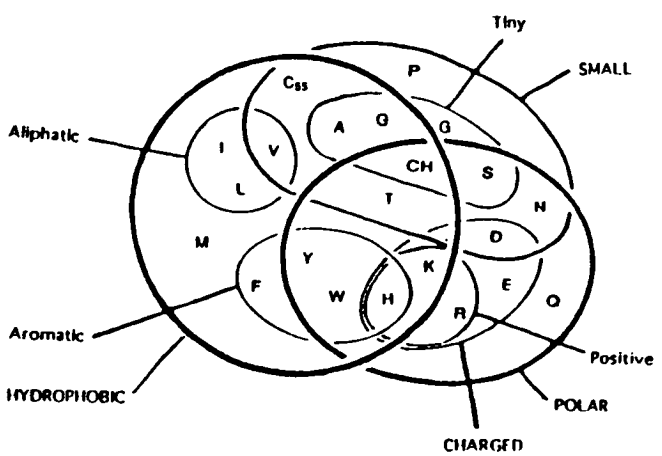


Figure 6.7. The Venn diagram shows the relationship of the 20 naturally occurring amino acids to a selection of physico-chemical properties that are important for the determination of tertiary structures of proteins (Taylor, 1986).

Gribskov (1987) has devised a method for searching protein databases for distantly related proteins or specific sequence motifs. The method uses a position-specific scoring matrix (profile), which expresses the information in a group of homologous sequences previously aligned by structural or sequence similarity. The positional scores are derived from some amino acid scoring matrix, which can be based on some quantifiable physicochemical property, but in the recommended application form, the 250 MDM odd log matrix of Dayhoff *et al.* (1978) is used (see appendix B). This means that each amino acid gets a specific score in the different positions of the protein sequence. This
score at a particular position is derived from some basic Dayhoff relatedness odd logs matrix and the amino acid composition at this position in the prealigned sequences. The same amino acid may score differently in different positions. The score reflects the likelihood for any amino acid to be found in a particular position in the protein structure.

In this discussion the profile method has been adapted for a graphical representation of sequence conservation to help clarify and analyse the primary structures of different cytochrome c sequence groups. It summarises and makes the information in the alignments more accessible for interpretation and enables a visual comparison of sequence characteristics between different subclasses of proteins of the same family, or between sequence motifs in apparently unrelated protein families.

In figure 6.8. the score values of the consensus sequences of the BacI cytochromes and ID subclass cytochromes from the purple bacteria, are plotted from their respective profile matrices. The graphs show where there is a pronounced conservation of a specific physicochemical property in each of the subclasses. On the same figure is plotted the score between the two subclasses It shows if a conservation of a particular physicochemical property "crosses the subclass boundary". In respect to the discussion on the possible relatedness of the BacI to the algal cytochromes f, the profile scores of BacI to the ID cytochromes and algal soluble f cytochromes are plotted together for comparison in figure 6.5a.

The information revealed in the profile analysis is as follows:

1) The different groups have a common conserved hydrophobic pattern (positions) in the areas which correspond to the alpha helix regions of the *Pseudomonas aeruginosa* cytochrome c551 structure (figures 6.8. and 6.11). The positions in between are much less conserved. In this context, it is

interesting to compare the region between the haem site cysteines and the region between the hydrophobic positions 59 and 63. While the residues between the cysteines are not especially well conserved in either subclass, the same or similar amino acids are used in both subclasses. In contrast, the residues after the sixth ligand methionine are well conserved in each subclass, but the chemical character of the conserved residues is different between the subclasses.

2) The least similar region between the subclasses appears to be right after the haem signature sequence C–CH. They are, however, highly conserved within each subclass and more so in the *Bacillus* subclass. This indicates that these regions fold differently in the different subclasses. There is indeed a sequence position in this region (position 19), which appears to be conserved. The amino acids in this position, in both subclasses, appear to be small and polar/acidic (by reference to Taylors' Venn diagram, see figure 6.7). The great majority of residues in this position have an oxygen atom on the terminal sidechain carbon atom, which may imply a functionally significant hydrogen bonding role. An amino acid of this type is indeed found, in this position, in other cytochrome subclasses, but this may be a fortuitous occurrence. Another region badly conserved is right after the sixth ligand met- pro dipeptide motif and involves three amino acids. This is a small loop region that apparently has different conformation in the two proteins.

6.4.2. Secondary Structure Analysis of P2-c552

A high resolution three dimensional structure is available for the *Pseudomonas aeruginosa* cytochrome c551, which belongs to the ID subclass. As the primary structure of P1-c552 has a considerable sequence similarity to the ID class cytochromes c, it is interesting to see if secondary structure predictions for the P1-c552 sequence correspond to that which is known about the *Pseudomonas aeruginosa* structure.

Secondary prediction methods currently have an accuracy limit at 65% for correct prediction (Garnier and Robson, 1989). Thus, there will always be an element of uncertainty concerning the results. To increase the credibility different methods can be used and/or secondary structure prediction of homologous sequences can be made. Congruence between the methods and/or between the sequences will strengthen the prediction.

Predictions were made by these two methods for three BacI sequences Bacillus azotoformans P1-c552, *Bacillus subtilis* c550 and *Bacillus licheniformis* c552 (figure 6.9.). At least four predictions out of six, indicate that three regions, corresponding to the first, third and the fourth helix regions of *Pseudomonas aeruginosa* cytochrome c551, similarly have alpha helix conformations. The region corresponding to the second helix region in the *Pseudomonas aeruginosa* structure has not as pronounced alpha helix character, despite considerable sequence similarity to the ID sequences. Still, three out of six predictions indicate such helix.

The alpha amphipathicity plots (Eisenberg *et al.*, 1989) of the *Ps. aeruginosa* c551 and P1-c552 cytochromes are similar (figure 6.10.). Noteworthy, is the pronounced alpha amphipathic character of the regions, which correspond to the second and the third helix regions in the *Pseudomonas aeruginosa* cytochrome. Interestingly they are also similarly delimited with a deep 'cleft' between the peaks. In *Pseudomonas aeruginosa* cytochrome c551 the sequence part corresponding to the cleft forms a small loop connecting the second and the third helices past the 'back edge´ of the haem.

6.4.3. Protein Modelling Studies on P1-c552. A Proposal for a Three Dimensional Structure for The P1-c552 Cytochrome C from *Bacillus azotoformans*

The similarity of *Bacillus azotoformans* cytochrome P1-c552 sequence and that of the *Pseudomonas aeruginosa* cytochrome c551 was deemed sufficiently extensive, for the former to be modelled from the latter. The computer modelling program Sibyl was used to enable visualisation of the structure. Proposals are made on how the insertion and deletions in the *B. azotoformans* structure may be accommodated and how the hydrophobic residues pack around the haem. Possible functional significance of some of the structural characteristics of the P1-c552 cytochrome is contemplated.

Protein modelling procedure

The modelling software Sibyl (Tripos Associates Inc, St Lous, USA)was used for the protein modelling; it has the following attributes

1) Builds molecules atom by atom

2) Builds polypeptides amino acid by amino acid

3) Secondary predictions of structures of polypeptides

4) Residue substitutions (mutations).

5) Modification of protein structures on the basis of hypothesis based on secondary predictions.

6) Loop search based on homology to a proposed loop region in the protein which is being modelled, the residues bordering the region being fixed.

7) Energy minimisation for parts or for the complete molecule.

The coordinate data (Matsuura *et al.*, 1982) of the *Pseudomonas aeruginosa* cytochrome c551 was obtained from the Brookhaven Protein Data bank. A program called Maxmin2, a part of the protein modelling software packet, was used for energy minimisation. The covalent or external energy terms included bond energies, bond angle energy, dihedral energy, and improper torsion energy. Noncovalent or external energy terms included electrostatic and van der Waals interaction energy. Hydrogen bonding was included implicitly in the electrostatic and van der Waals terms.

A stepwise procedure of modelling of the P1-c552 cytochrome was followed. First, all the surface residues were substituted with corresponding amino acids in the P1-c552 cytochrome sequence. It was reasoned that they cause a minimal disturbance in the three dimensional structure. The obtained structure was then energy minimised. The interior residues were then changed one at a time and the resultant structure subjected to local energy minimisation before another proposed residue was changed. The regions with no corresponding sequence part in the *B. azotoformans* sequence were then deleted and the bordering atoms joined. An energy minimisation was then applied and the region further modelled, if possible, on a loop of a maximum homology.

6.4.3.1. The basic backbone structure of P1-c552

Positions and structural components (secondary structural elements) are numbered and defined according to the alignment of the *Bacillus azotoformans* cytochrome P1-c552 and *Pseudomonas aeruginosa* cytochrome c551 structure in figure 6.11,. Residue numbers will refer to positions in this alignment, not to the actual sequence number. Spatial designations will refer to the orientation in the schematic three dimensional picture of *Pseudomonas aeruginosa* cytochrome c551 in

figure 6.12, (frontal view onto the edge of the haem with the propionates of the haem pointing downwards).
In the following sections, the two structures are compared region by region as they are depicted in the primary structure of the *Pseudomonas aeruginosa* cytochrome c551 in figure 6.11. Figure 6.13a shows the computer model of alpha-carbon backbone chain of the *Ps..aeruginosa* cytochrome c551 drawn from coordinates deposited in The Brookhaven Databank (Matsuura *et al.*, 1982), figure 6.13c shows the alpha-carbon backbone chain of the *B. azotoformans* cytochrome P1-c552 modelled on the former and figure 6.13b shows the structures superposed

6.4.3.2. The N- and C-terminal helices

The proteins begin and end in alpha helices that interact with each other. In the N-terminus helix region the main difference between the structures is deletion of a near invariant lysine in position 10 in P1-c552. This lysine forms the terminal part of the helix in *Pseudomonas aeruginosa* cytochrome c551, and may be a specific insertion, characteristic to the ID subclass cytochromes in purple bacteria. The configuration without lysine as in *B. azotoformans* cytochrome c552, is found in other subclasses and may be a more ancient trait (Moore and Pettigrew, 1990).

6.4.3.3. The loop region, Cys 15 to Pro 26

The uncertainty in the model is greatest just after the haem site cysteine (position 15). This had indeed been anticipated from the profile analysis. The region is anchored at the ends, by the haem site cysteine and the histidine (positions 15 and 16) at one end and the proline, which hydrogen bonds to the fifth ligand histidine, at the other end. The sequence part preceding the proline forms a bow like structure that turns upwards and the glycine (position 25) and the valine (position 24) form hydrogen bonds with cysteine (position 15). This bow-like conformation is found in the corresponding region of other cytochromes belonging to different subclasses, such as *Rhodospirillum rubrum* cytochrome c2, Tuna cytochrome c and *Azotobacter vinelandii* cytochrome c5. It is not found in *Chlorobium limicola* c555, where the corresponding region extends downwards. The bow like structure may be an ancient near invariant structural feature.

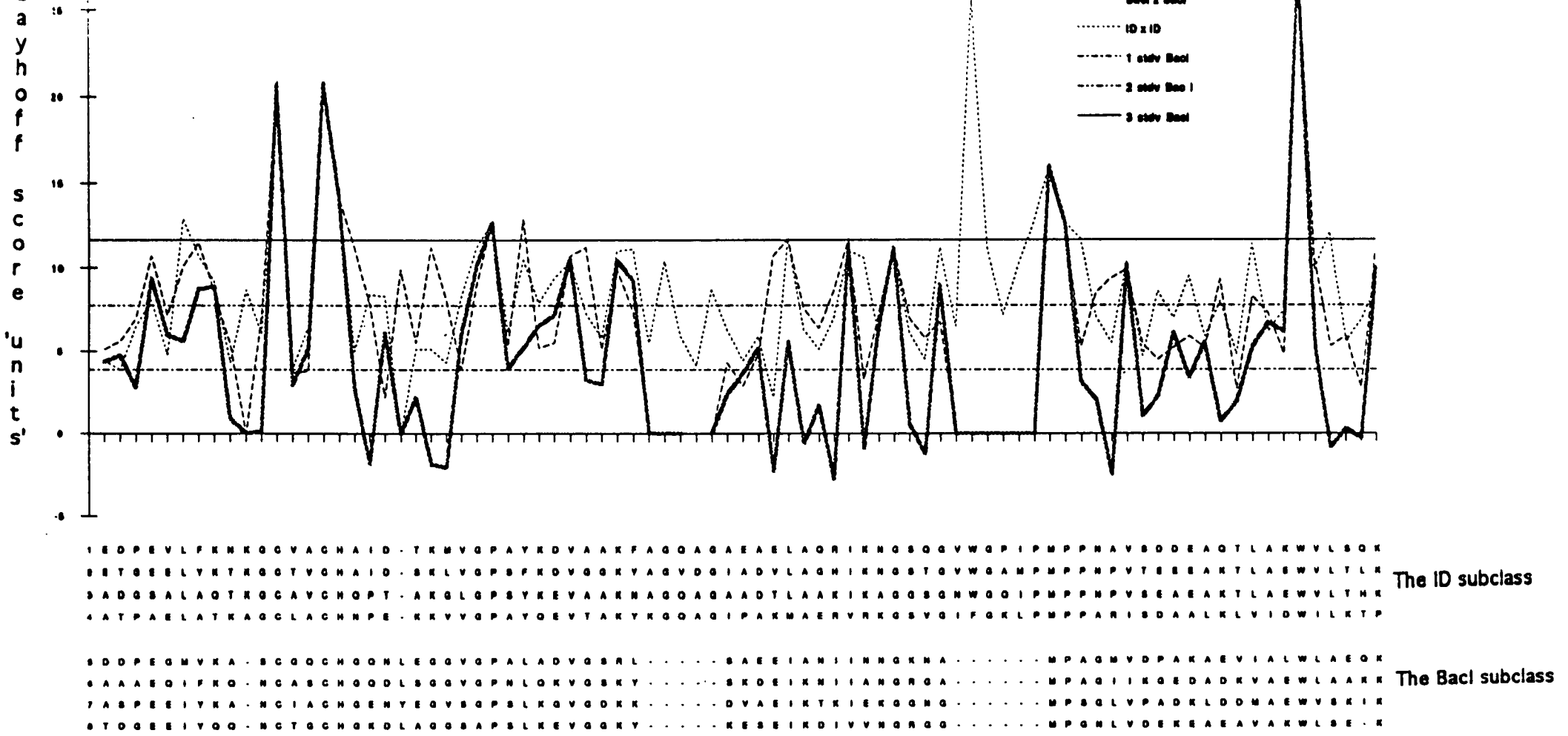


Figure 7.8. Plots of the highest Dayhoff scores/sequence position in the profile matrices of the BacI- and ID subclass sequences and the score (ID x BacI) between them.

The ID x BacI score reflects the positional conservation between the subclasses. Three guidelines are inserted, one times, two times and three times the standard deviation from the mean score in the profile matrix of the BacI cytochromes c. The profiles are derived from the Dayhoff 20 x 20 amino acid score matrix (PAM 100). The ID profile was calculated from 14 sequences four of which are shown (only common core elements). They are from: 1: *Ps. aeruginosa*, (Ambler 1963); 2: *Azotobacter vinelandii*, Ambler 1991); 3: *Pseudomonas fluorescens*, (Ambler, 1991); 4: *Rhospseudomonas gelatinosa*, (Ambler 1991) and 4: 4: The BacI is based on four sequences. They are from: 5: *Bacillus azotoformans* (this work); 6: *Bacillus* PS3, (Fujiwara etal., 1991) 7: *Bacillus subtilis*, (Von Wachenfeldt et al., 1991) and 8: *Bacillus licheniformis*, (Woolley (1986) and VanBeeuman personal communication).

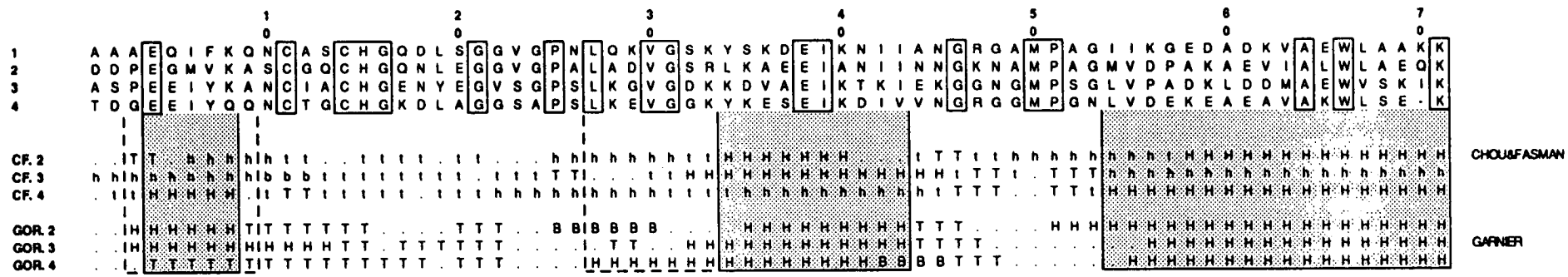
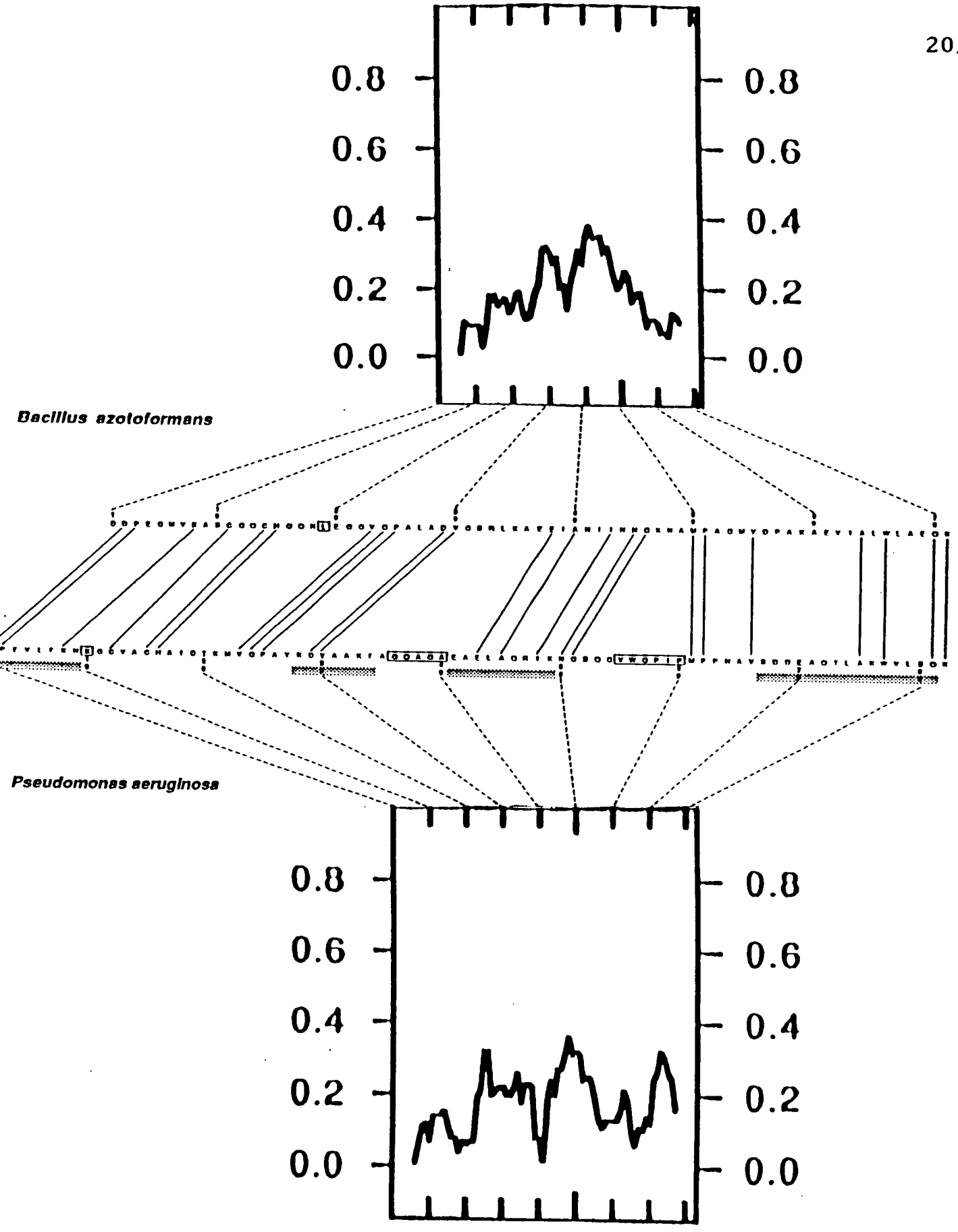


Figure 6.9. Secondary structure prediction of cytochromes c of the Bac I subclass.

Secondary structure predictions were made according to the methods of Chou & Fasman (Fasman,1989) and Garnier & Osguthorpe &Robson (Garnier and Robson 1989). These methods were applied to three of the four cytochrome c sequences of the BacI subclass (as it is defined in this thesis) in the above alignment. The pairwise alignments have about 40% identical residues among each other, overall, 18 of the 71 residues are conserved in all sequences. to CF.2-4 are Chou-Fasman predictions, GOR. 2-4 are Garnier-Osguthorpe-Robson predictions. h=a weak helix, H = strong helix, b = a weak beta, B = a strong beta, t = a weak turn, T = a strong turn, . = other.

1 ***Bacillus*** PS3 (Fujiwara et al. 1993)
2 ***Bacillus azotoformans*** P1-c552 (Hreggvidsson this work).
3 ***Bacillus subtilis*** c550 (VonWachenfeldt et al., 1990).
4 ***Bacillus licheniformis*** c551 (Wolley (1986) and VanBeeuman personal communication).



re 6.10. Alpha amphicity plots of the cytochrome P1-c552 from *Bacillus azotoformans* and the cytochrome c551 from *Pseudomonas aeruginosa* (Eisenberg, 1989).

□ = Insertion ▒ = Alpha helix regions in the *Pseudomonas aeruginosa* cytochrome c551.

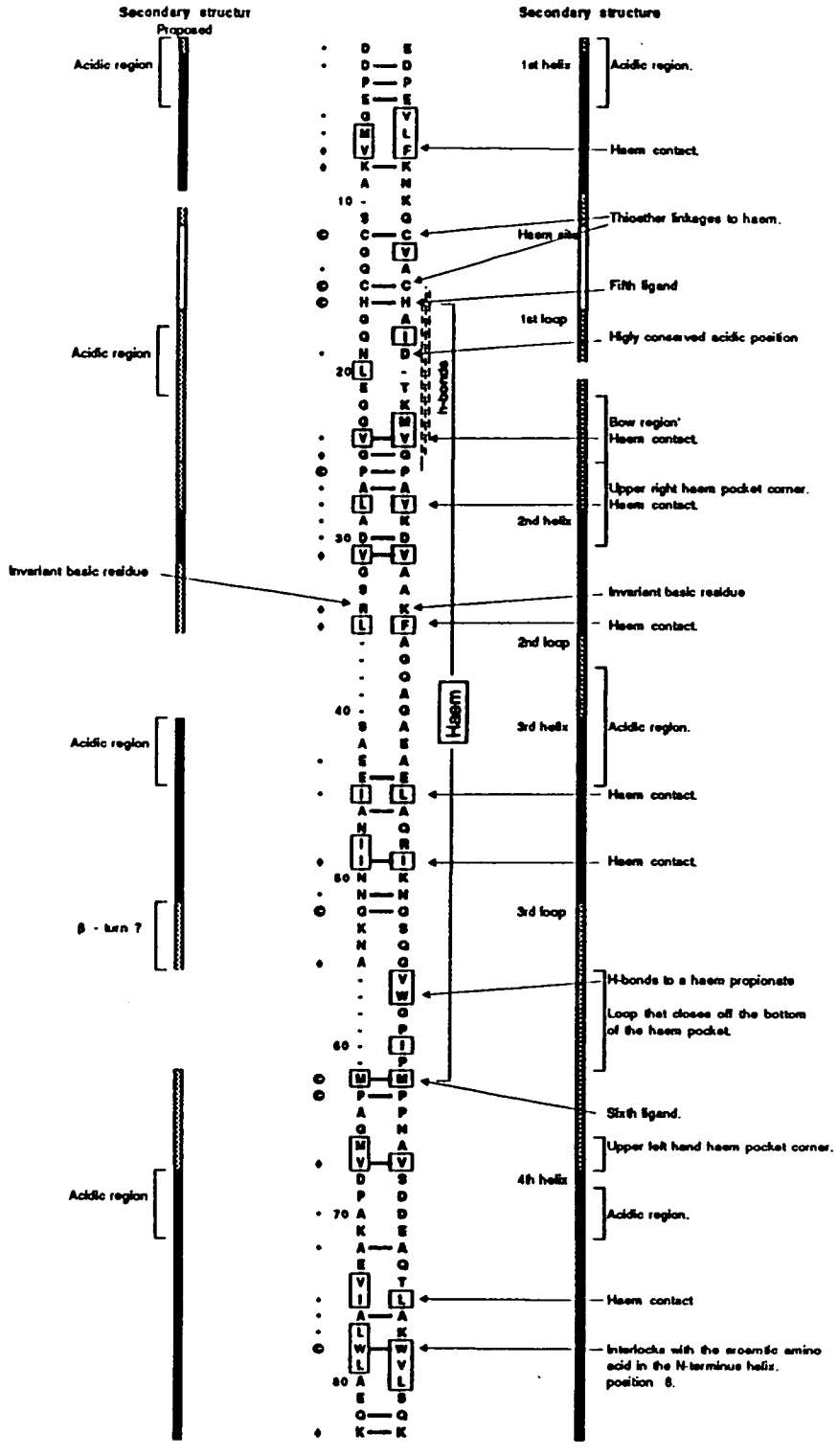


Figure 6.11. **An alignment of the cytochrome c P552 from *Bacillus azotoforman* and the cytochrome c551 from *Pseudomonas aeruginosa* A structural comparision based on primary sequenc characteristics.**

Positional amino acid conservation between the corresponding subclasses, ID and Bacl, is estimated on the basis of profile analysis represented in figure 6.8.

- ⊙ - Invariant between the subclasses
- ♦ - Conserved between the subclasses. The positional ID x Bacl score is > 2 x stdv above the mean score in the Bacl profile matrix.
- • - Semiconserved' between the subclasses. The positional ID x Bacl score is < 2 x stdv, and > 1 x stdv above the masan score in the Bacl subclass.

The secondary structure poposal for the P1-c552 cytochrome is based on the observed affinity of the primary structrure to the one of the cytochrome c551 from *Pseudomonas aeruginosa*, of which the three dimensional structure is known. It is also based on the secondary structure predictions (figure 6.9 and 6.10).

- An alpha helix conformation
- A loop
- The haem site sequence part.

[X] Hydrophobic amino acids in the sequences are boxed. Acidic region are delimited on the basis of the subclass alignments in figure 6.5a.

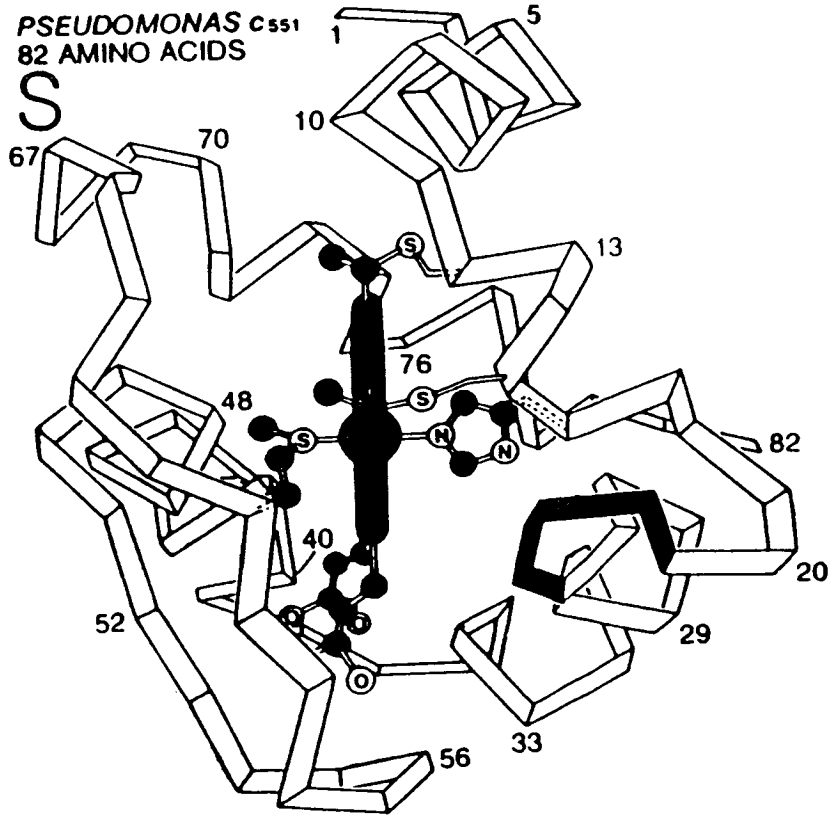


Figure 6.12. A schematic representation of the three dimensional crystal structure of *Pseudomonas aeruginosa* (from Dickerson, 1980).

Residue numbers refer to figure 6.10.

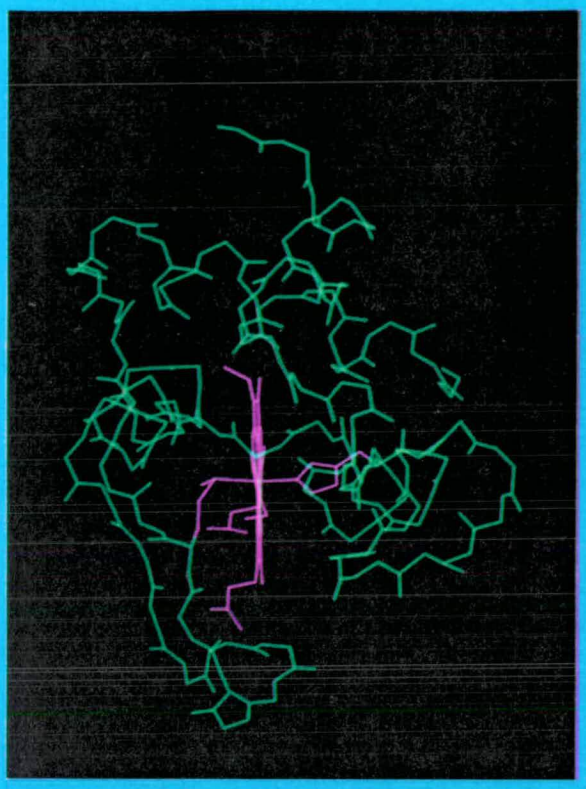


Figure 6.13a.

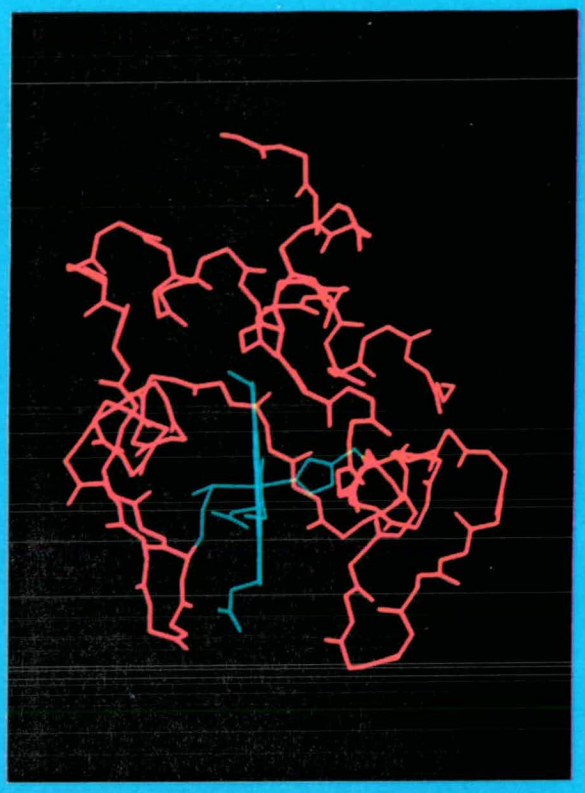


Figure 6.13b.

Figure 6.13c.

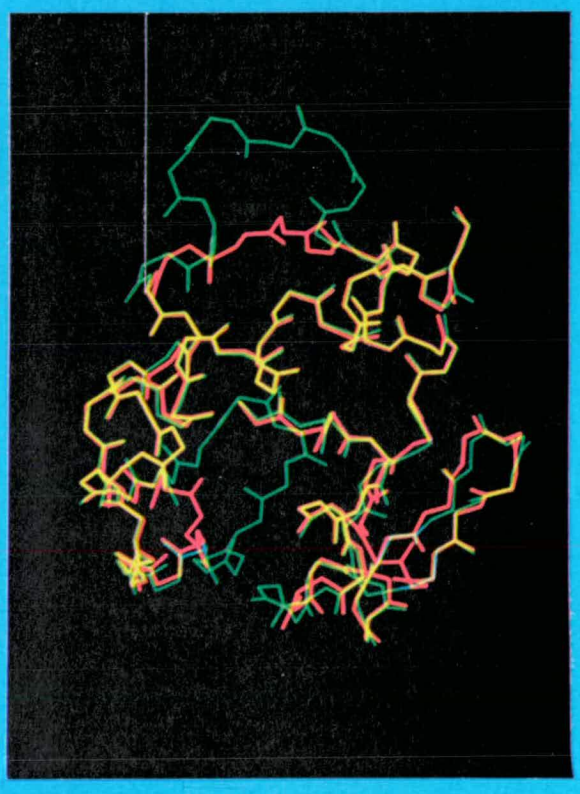


Figure 6.13d.

Figure 6.13. Computer models of the alpha-carbon backbone chains of *Pseudomonas aeruginosa* cytochrome c551 and *Bacillus azotoformans* cytochrome P1-c552.

The c551 cytochrome is drawn from coordinates from the Brookhaven Databank. The P2-c552 cytochrome is modelled on that structure The structures (except in figure 6.13d) are viewed from the front with the propionates pointing downwards.

Figure 6.13a. *Ps. aeruginosa* c551

Figure 6.13b. *B. azotoformans* P1-c552 .

Figure 6.13c. *Ps. aeruginosa* c551 anc *B. azotoformans* P1-c552 superposed.

Figure 6.13d. *Ps. aeruginosa* c551 anc *B. azotoformans* P1-c552 superposed. The structures are tilted so that the c551 'backside' loop is exposed.

2) Builds polypeptides amino acid by amino acidLoop search was performed for this region, but none came up which satisfied the condition that the hydrophobic residue in position 20 is buried. In all cases the sidechain protruded into the surrounding. This is unlikely, especially, if one considers that this residue is tyrosine in the homologous sequence of *B. licheniformis* cytochrome c552 (figure 6.15b). The presence of glycine in position 17 and the pair of glycine at the other end, in positions 22 and 23, may enable the chain between, to take up a specific conformation that docks the hydrophobic residue, in position 20, into the groove between the end of the C terminal helix and the cysteines.

The similarity of this region to the comparable region of c4 cytochromes and the P3&P4-c555 has been pointed out in section 6.3.3.

6.4.3.4. The helix region, Pro 26 - Lys 34

The second helical region in the *Ps. aeruginosa* c551 cytochrome has a close counterpart in the P2-c552 cytochrome. All the substitutions are conserved, the same hydrophobic pattern is observed and there is a strictly invariant basic residue in position 34 in both sequences. The hydrophobic residue (sometimes bulky and aromatic) of the adjacent residue (position 35) is also conserved. The 26-pro-ala-tyr-28 sequence forms the upper right hand corner of the haem crevice in *Ps. aeruginosa* c551. The corresponding tripeptide, 26-pro-ala-leu,-28, is likely to do the same in the P1-c552 structure. These parts of the different sequences may therefore have the same conformation, that of an alpha helix. However, the strict conservation of glycine, in position 32 in the P1-c522 sequence, perhaps indicates that the polypeptide backbone chain from there on, may have a different conformation (see also the next section).

6.4.3.5. The loop region, Gly 35 to Gly 40

The region from second helix to the third helix, is a small loop bordered in by glycines in positions 36 and 40 in the *Pseudomonas aeruginosa* c551 cytochrome. The glycines are well conserved and appear to make the chain loop outwards (figure 6.13d.), However, they are not invariant as they are not found in at least two ID subclass sequences, those of *Rc. purpurus* and *Ps. hydrogenothermophilus.* (figure 6.5). The loop seems to be deleted in the P1-c552 structure (figure 6.13.d). Therefore amino acids, which in *Ps. aeruginosa* structure have an alpha helix conformation, must take up a more extended conformation in the P1-c552 structure to span the gap between the second and the third helices in *B. azotoformans,*.

In position 32 in the P1-c552 structure, there appears to be a strict requirement for glycine. According to Richardson and Richarson (1989) 'one would expect glycine to occur where there is a special need for its presence'. Glycine has a much greater conformational freedom than other amino acids. It is also the smallest one. It is common in turns and it occurs where there is a special need for a small amino acid, as in the middle of alpha helices where they interact. However, it is also a good alpha helix breaker. In its position (32) in the P1-c552 structure, it may be needed for either of the following reasons. Either its smallness may allow the second helix structure to pack closer to the haem or it may indicate that the backbone takes up a different conformation from the helicitity observed in the *Ps. aeruginosa* structure. The glycine would then act as a breaker residue to bend the chain behind the haem in a similar way to that observed in the bacterial c2 sequences (from glycine in position 34 in the Tuna cytochrome c in figure 6.18b and position 35 in the *Rh. rubrum* cytochrome c2 in figure 6.18c). A glycine is highly conserved in this same position in the bacterial c2 sequences (figures 6.3 and 6.17a).

6.4.3.6. The third helix region.

According to the secondary prediction studies above, residues from alanine (42) to aspargine (50) form a helix. It may be an equivalent structure to the *Pseudomonas aeruginosa* cytochrome c551 third helix. Position 41 in the *Bacillus* sequences has a greater number of N-cap residues than other positions in the region. It may indicate the start of the third helix in the *Bacillus azotoformans* sequence. The hydrophobic pattern is identical to the one in the *Pseudomonas aeruginosa* cytochrome c551 third helix, but the residues in between are not as well conserved.

6.4.3.7. Third loop, from position 50 to the sixth ligand methionine in position 62.

In this region the *Pseudomonas aeruginosa* sequence is longer by six amino acids. This part of the molecule forms a loop which partly closes the haem crevice and shields the propionates to some extent. A conserved tryptophan is found in this loop, which it is proposed hydrogen bonds to one of the propionates. In the *Bacillus azotoformans* sequence the sequence is just long enough, five residues, to be able to break from the preceding alpha-helix and connect the sixth ligand to the haem. This requires a tight turn, and the amino acids are typical of β -turns including

at the start a good alpha helix breaker in an aspargine (Richardson and Richarson, 1989).

6.4.3.8. The fourth helix region from the sixth ligand methionine, in position 62, to the end, position 82

According to the profile analyses the region is not particularly well conserved. However the hydrophobic pattern is conserved and the tryptophan in position 78, a characteristic feature of the ID subclass, is fully conserved (figure 6.12). No gaps are found. Apart from a stretch of 2-4 amino acids following the sixth ligand methionine, secondary predictions (figure 6.9) strongly indicate an alpha helix conformation of this region, which is the conformation in the *Ps. aeruginosa* structure.

6.4.4. The Distribution of Charged Amino Acids in The P1-c552 Cytochrome

In figure 6.5. which shows the proposed alignment of the BacI sequences to the ID subclass sequences and the algal cytochrome c sequences, the positive residues are marked with blue and the negative ones with red. Apparently, the charged residues cluster in several patches and positions, with a predominance of either positive or negative residues in each. The distribution of charged amino acids is noticeably more different between the algal and the BacI sequences than between the BacI and the ID subclass sequences. However, the positional variation is much greater between sequences in the BacI subclass than in the ID subclass, even though the latter subclass consists of sequences from many different genera and the former of sequences from only one genus. It has been proposed that charged amino acids contribute in determining the specificity of cytochromes c towards reactivity to other redox proteins (Salemme *et al.*, 1973, Salemme, 1977). The observed diversity in their placement in the different BacI sequences, may be explained by the presence of the two BacI cytochromes in *Bacillus* species. They may have different roles in the electron transport pathway, with a consequent selection for divergence.

While the conservation of charged amino acids is not positionally as distinct in the BacI sequences as in the ID sequences, the patches generally correspond regionally within and between the subclasses. In the first helix the distribution is almost identical, but for the loss of the

lysine residue in position 10 in the BacI sequences. There is a strict conservation of lysine in position 22 in the first loop of the ID subclass sequences, but it is not found in the corresponding loop of the BacI sequences. This is a further indication of the structural differences of these regions between the subclasses.

The third helix regions in both subclasses start with a conserved patch of acidic residues, which is more pronounced in the BacI cytochromes. The remaining part has a pattern of predominantly basic residues, but which is not the same in the different sequence groups. Furthermore, a requirement for positional conservation appears to be stricter in the ID subclass.

Acidic residues, in the C-terminal helices of the BacI sequences, are not as strictly confined to the so-called acidic region as in the ID cytochromes. This acidic region is a tripeptide sequence, from position 69 to 71 in the ID subclass and it is highly conserved in most other subclasses as well. In the BacI sequences, acidic residues are found in a greater number of positions in the C-terminal helix, than in the C-terminal helices of the ID sequences. These acidic residues, interspersed with single basic residues are found predominantly at the start of the helix and include the tripeptide acidic region.

In figure 6.14a,b,c, and d, the distribution of charged amino acids in the BacI cytochromes, from *B. azotoformans* , *B. licheniformis* and *B. subtilis* are compared to the *Pseudomonas aeruginosa* Approximately the same distribution pattern of the charged patches on the surface, is observed. While the basic amino acids are more or less evenly distributed (figure 6.14a,b,c,d), there is a greater proportion of acidic amino acids placed peripherally on the upper left side, at the back and the top in other BacI cytochromes. It has been proposed that such an asymmetrical charge distribution helps to orientate the cytochrome towards the membrane, cell wall and to its electron donors and/or acceptors (Salemme, 1977).
The edge of the haem is partly exposed at the front in cytochromes c. This exposure is only about 4% in mitochondrial cytochromes (Stellwagen 1978, cited in Moore and Pettigrew, (1990). In contrast to mitochondrial cytochromes, which have lysine residues on the front face close to the edge of the haem, all residues next to the frontal edge of the haem in the ID subclass cytochromes are uncharged. Both types share, however, a ring of lysines around the periphery of the front face (Moore and Pettigrew , 1990). It has been suggested that the peripheral lysine residues may be important for enabling and defining the specificity of the interaction

with other redox proteins (Salemme *et al.*, 1973; Salemme, 1977). The BacI cytochromes c appear to be similar to the ID cytochromes in this respect (figures 6.14 and 6.16). They have uncharged residues at the border of the haem crevice, but from two to four peripheral lysines in the same or analogous positions to that observed in *Pseudomonas aeruginosa* c551. Two of these lysines, in positions 8 and 28 (numbers refer to the *Ps. aeruginosa* c551 structure figure 6.12) are strictly or highly conserved in both the subclasses. The others are not conserved in the BacI cytochromes, but lysines in different positions and approximately in the same regions around the periphery, may serve the same purpose. The lysines in the upper left hand corner, in any of three positions, 67, 69 and 70, in the PS3, *B. azotoformans*, *B. subtilis* and *B. licheniformis* cytochromes, would then be analogous to the peripheral lysine, in position 10, in the ID sequences. These particular lysines may make salt bridges or hydrogen bond with acidic residues on the N-terminal helix. Furthermore, there are additional peripheral lysines, in positions 18 and 50, in the *B. subtilis* 550 sequence.

Unusually few basic amino acids are found in the *B. azotoformans* P1-c552 cytochrome. They may therefore be the functionally most important ones. Of these, only two are peripheral, the lysines in positions 8 and 70. They correspond to the *Pseudomonas* c551 lysines, in position 8 and 10, respectively. The other three basic residues are strictly or highly conserved. One is a lysine at the end of the sequences, but lysines are commonly found at the C-terminus in other cytochromes as well as in other proteins. Another basic residue, an arginine, is found in the second helix region in position 33. It aligns with an invariant lysine residue in the ID subclass sequences and a lysine is found in this position in the other BacI sequences. Its role has not been guessed at in the ID sequences, but perhaps it hydrogen bonds to one of the propionates. Such may also be the role of the lysine or the arginine that is found in the proposed left side β-turn, in position 53, in *Bacillus* PS3 c551, *Bacillus azotoformans* P1-c552 and *Bacillus licheniformis* c552 sequences (figures 6.11, 6.14b,c).

If the model for the P1-c552 cytochrome is right, the propionates are exposed to the solvent. All the BacI sequences have at least two lysines on the chain where it borders the opening any of which may bond to the propionates. These lysines in the different structures are in positions 33, 36, 34, 46, and 53. These are superposed on to the *B. azotoformans* P2-c552 structure in figure 6.14e.

6.4.5. An Assessment of The P1-c552 Protein Model

On the basis of extensive structural comparisons of proteins in different families, Chothia and Lesk (1986) have observed that the degree of success expected in predicting the structure of a protein from its sequence, using the known structure of a homologous protein, depends on the extent of the sequence identity. Thus, a protein will provide a close general model for other proteins if its sequence homology is larger than 50%. If the identity drops to 20% there will be large structural differences, that are at present impossible to predict (Chothia and Lesk, 1986).

The identity between the P1-c552 from *Bacillus azotoformans* and *Ps. aeruginosa* c551 cytochrome is only about 37% and therefore some structural differences may be expected. On the other hand the size of the P1-c552 structure is small, there are no loops to speak of, so that the sequence has relatively fewer configurations to choose from and therefore is perhaps easier to predict. The polypeptide chain has to wrap the haem and supply hydrophobic residues into essential spatial positions relative to the haem. There appears to be a particular or relatively strict requirement for a certain pattern of hydrophobic residues around the haem. Furthermore, even if the identity is minimal, the secondary structure prediction of the core regions indicates that they are similar between the proteins. Both the structures can also be anchored by certain residues, which are invariant between the protein subclasses. These are the cysteines that covalently bind the haem and the haem ligand residues, the histidine and the methionine. The proline in position 26, which hydrogen bonds with fifth ligand histidine, may also be considered an important anchoring point, as it has to be within a certain distance and orientation with respect to the histidine.

According to the above modelling, it appears that the P1-c552 protein consists essentially of only the core components of the *P. aeruginosa* structure. This structure is obtained by stripping the *P. aeruginosa* cytochrome c551 of all loops and leaving only the bare minimum of sequence for wrapping the haem (figure 6.13a,b,c,d). It should also be emphasised that there are regions that obviously are different and that even if the core elements are homologous, they can be spatially shifted relative to the haem in the different structures. These will never be realistically predicted at this stage by the available data. The energy minimisation program only operates by finding the local energy optimum and the configurations will therefore not stray very far from those

prescribed by the c551 structure, on which the model is based. The best or lowest energy solution to the three dimensional structure, can therefore easily be overlooked. Attempts at more refined structure modelling in the non homologous regions were deemed futile or still an exercise more properly in the domain of protein folding theorists, needing to be supported by experiments and crystallographic data.

6.4.5.1. Exposure of hydrophobic and charged residues

One way of evaluating model structures is to examine to what extent hydrophobic amino acids are buried and hydrophilic amino acids are accessible to the solvent. In figure 6.15a are shown only the hydrophobic amino acids of the P1-c552 structure and those in corresponding regions of the *Pseudomonas aeruginosa* cytochrome after superposition of the structures. In the *Bacillus* P2-c552 model all except one or maybe two, of the conserved hydrophobic amino acids are buried and those that have counterparts in the *Pseudomonas aeruginosa* c551 structure have similar side chain conformation. The most prominent exception is the leucine residue in position 20, which is in the apparently non homologous loop, between the haem site and the conserved proline site. The hydrophobic amino acid side chains in position 45 may be exposed to the solvent but the conserved isoleucine in position 47 appears to be buried. In figure 6.15b are shown the hydrophobic amino acids of the structures of the superposed *Pseudomonas aeruginosa* c551 structure and of three BacI models.

All the charged residues in the P1-c552 model occupy surface positions, except for the one in the proposed β-turn (section 6.4.3.7, figures 6.11 and 6.14b). This residue may hydrogen bond to one of the haem propionates. Apparently the lysine side chain lies towards the interior of the protein. Such a role is proposed for histidine in position 47, of the *Pseudomonas mendocina* ID subclass cytochrome c (Moore and Pettigrew, 1990).

6.4.5.2. Shielding of the haem

Any factor that helps to stabilise the oxidised form of a redox couple makes it a better electron donor with correspondingly lower redox potential. If the reduced form is stabilised, the redox couple is a better electron acceptor and has correspondingly higher redox potential (Moore and Pettigrew, 1990). The converse is true about destabilising factors of either of the redox states.

There are many factors that affect the redox potentials of cytochromes c, one of which is the degree of haem shielding. It is proposed that by increasing shielding the hydrophobicity around the haem is increased. This has a destabilising affect on the oxidised (higher charge) state therefore the redox potential may be expected to become more positive. Conversely, when the haem shielding is decreased the hydrophobicity decreases and the redox potential becomes less positive. The results of Schlauder and Kessner (1979, cited in Moore and Pettigrew (1990), support this. They observed, when comparing the close relatives, cytochrome c2 from *Rhodospirillum rubrum* and cytochrome c from horse heart, that the less positive redox potential is associated with a greater haem exposure. However, Moore and Pettigrew (1990) caution against too much generalisation. They emphasise that this may not hold when nonhomologous structures are compared and that redox potential may be determined by a complicated interplay of various factors. Thus, even though the *Prosthecochloris aestuarii* cytochrome c-555 has a lesser haem exposure than the horse cytochrome c, the former has a far less positive redox potential. The redox potentials of these cytochromes are, respectively, 102 mV and 203 mV, (Schlauder and Kassner, 1978 Fietchner *et al.*, 1978, 1979, cited in Moore and Pettigrew 1990).

The smaller BacI sequence may be expected to have greater haem exposure than cytochrome of the ID subclass. There is less protein to wrap the haem with, the haem crevice may be tighter (the missing backside loop) and the haem may therefore exert a greater strain on the backbone chain wrappings. Also the loop which partly closes off the haem crevice in *Ps. aeruginosa* c551 cytochrome, from position 53 to 59, is apparently missing in the *Bacillus* structure, and the propionates may be more extensively bared in the *B. azotoformans* structure (figure 6.13c). Due to this proposed larger degree of exposure of the haem, the oxidised state would be expected to be less destabilised in the BacI cytochromes than in the ID subclass sequences and the redox potential correspondingly lower. This indeed appears to be the case. The ID subclass sequences cluster in midpoint redox potential between 200 mV and 300 mV (Moore and Pettigrew, 1990), but the BacI cytochromes cluster around a significantly lower value, around 120 mV. The only exception is the *Bacillus* PS3 cytochrome which is measured to have a midpoint redox potential of 212 mV. This is surprising considering the similarity of that cytochrome to the partial sequence of *Bacillus halodenitrificans* (midpoint redox potential 130 mV) and there are also no obvious primary sequence features of the

PS3 cytochrome which may explain this difference. The cytochrome c553 from *Rhodocyclus purpureus* has a much higher midpoint redox potential than the other ID cytochromes. It has a few amino acid insertion in the sequence region which closes off the haem, which indicates that the haem is more shielded. This would be in accordance with the theory that greater haem shielding results in higher midpoint redox potential.

Figure 6.14a.

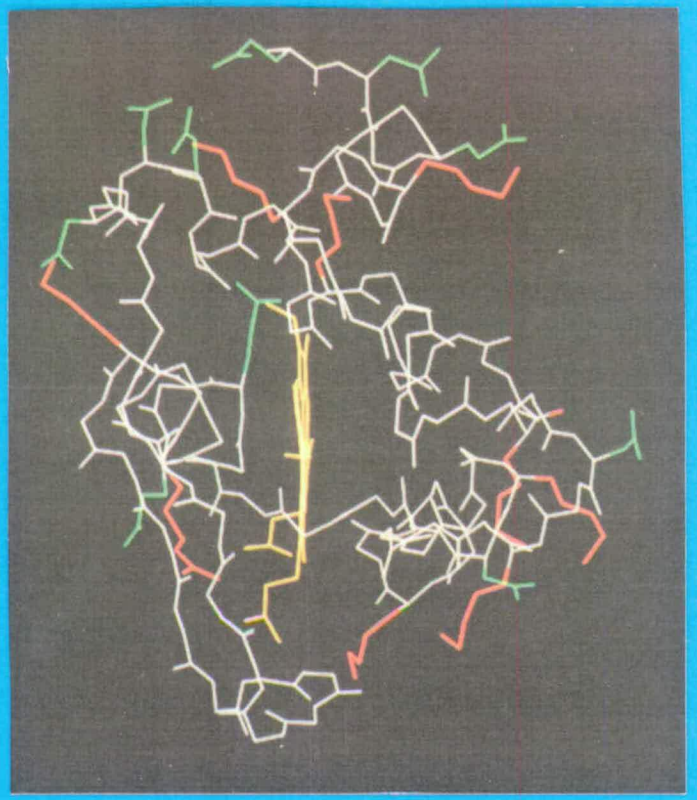


Figure 6.14b.

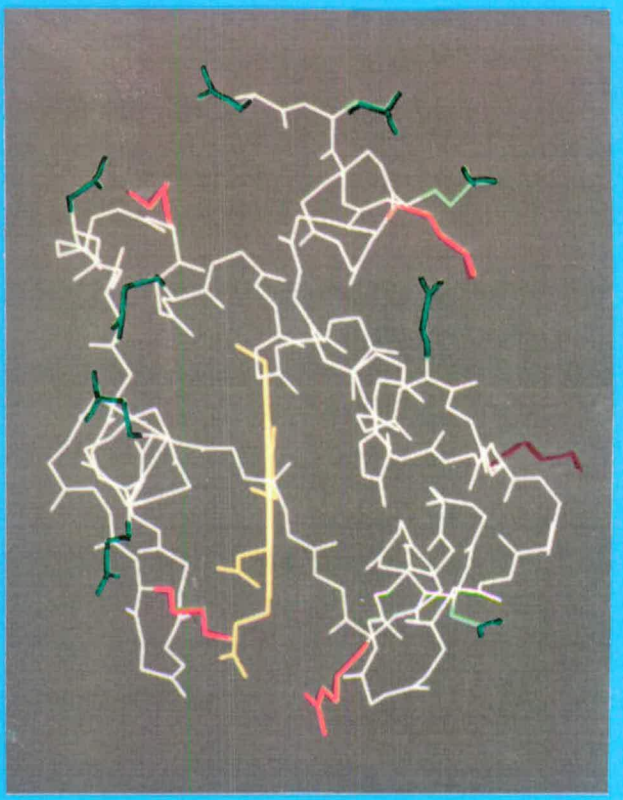


Figure 6.14. Computer models of *Pseudomonas aeruginosa* cytochrome c551, *B. azotoformans* cytochrome P2-c552, *B. licheniformis* cytochrome c552 and *B. subtilis* cytochrome c550, showing the distribution of charged amino acids.

The structures are viewed from the front with the haem propionates pointing downwards.
Sidechains coloured green = negtive amino acids, glutamate and aspartate.
Sidechains coloured red = positive amino acids, arginine and lysine.

Figure 6.14a: *Pseudomonas aeruginosa* c551
Figure 6.14b: *Bacillus azotoformans* P2-c552
Figure 6.14c: *Bacillus licheniformis* c552
Figure 6.14d: *Bacillus subtilis* c550
Figure 6.14e: Positive amino acids in the three *Bacillus* sequences borderingthe bottom of the haem superposed onto the *Bacillus azotoforomans* P2-c552 structure

Figure 6.14c.

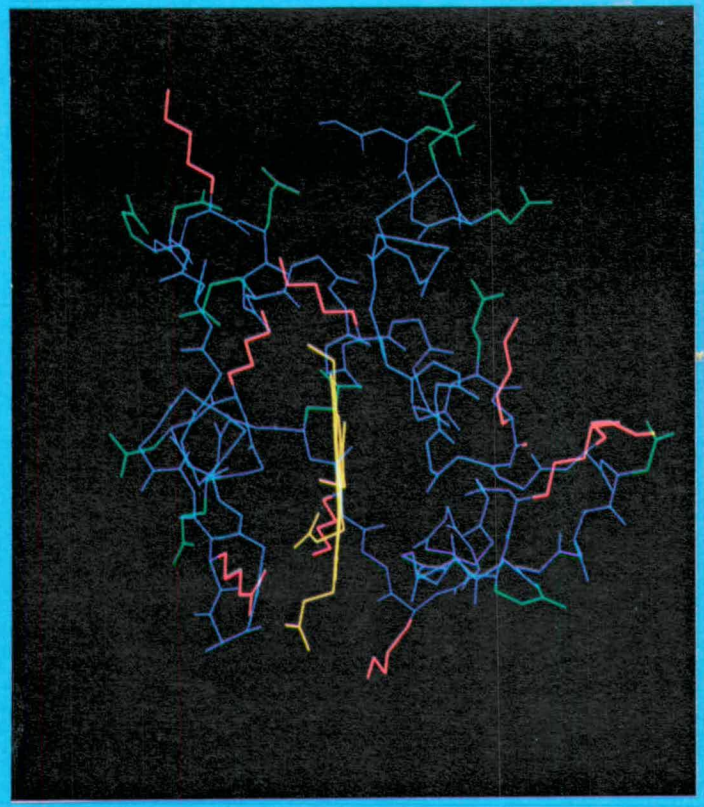


Figure 6.14d.

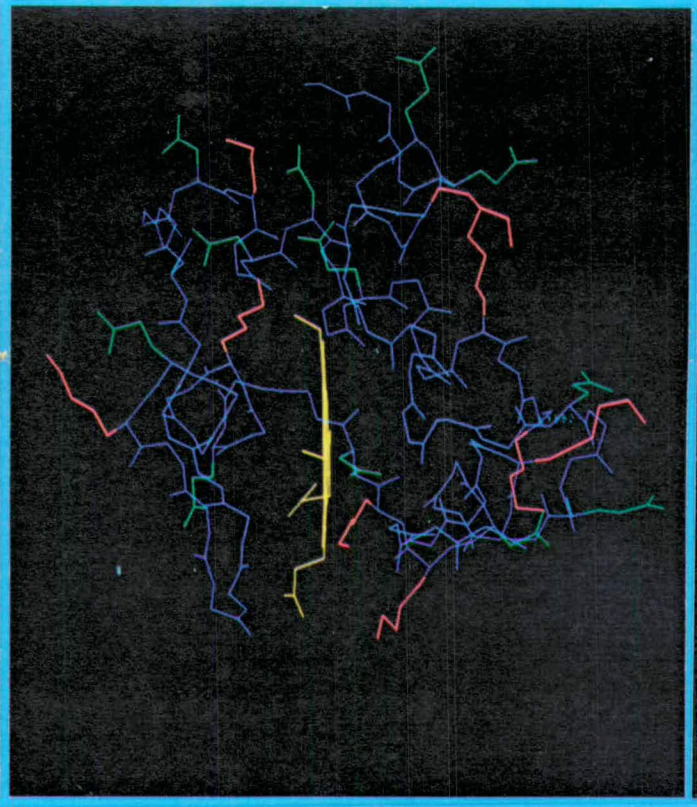


Figure 6.14e.

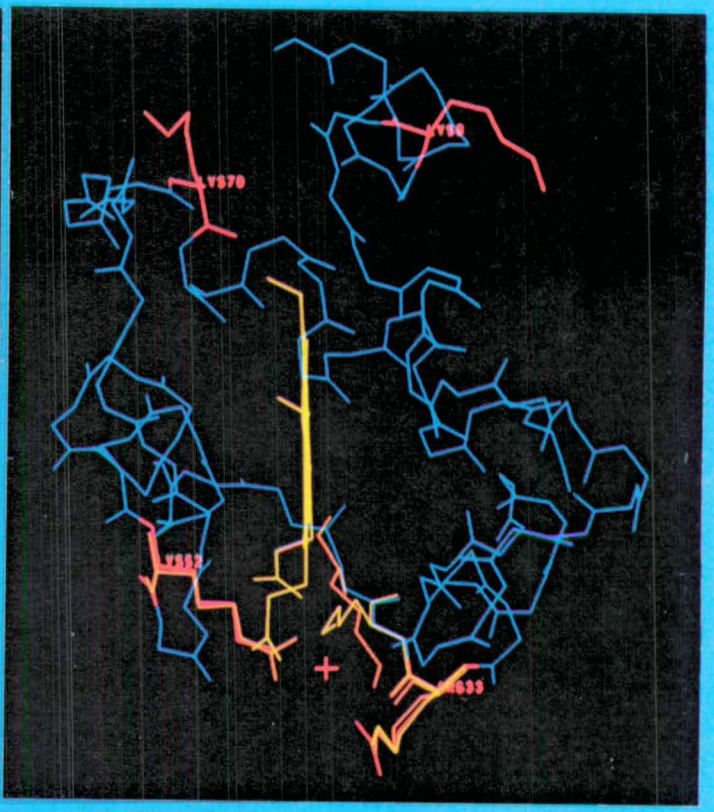


Figure 6.14 (continu.). Computer models of *Pseudomonas aeruginosa* cytochrome c551, *B. azotoformans* cytochrome P2-c552, *B. licheniformis* cytochrome c552 and *B. subtilis* cytochrome c550, showing the distribution of charged amino acids.

The structures are viewed from the front with the haem propionates pointing downwards.
Sidechains coloured green = negtive amino acids, glutamate and aspartate.
Sidechains coloured red = positive amino acids, arginine and lysine.

Figure 6.14a: *Pseudomonas aeruginosa* c551
Figure 6.14b: *Bacillus azotoformans* P2-c552
Figure 6.14c: *Bacillus licheniformis* c552
Figure 6.14d: *Bacillus subtilis* c550
Figure 6.14e: The positive amino acids, of the three *Bacillus* sequences above, bordering the bottom of the haem superposed onto the

Figure 6.15a

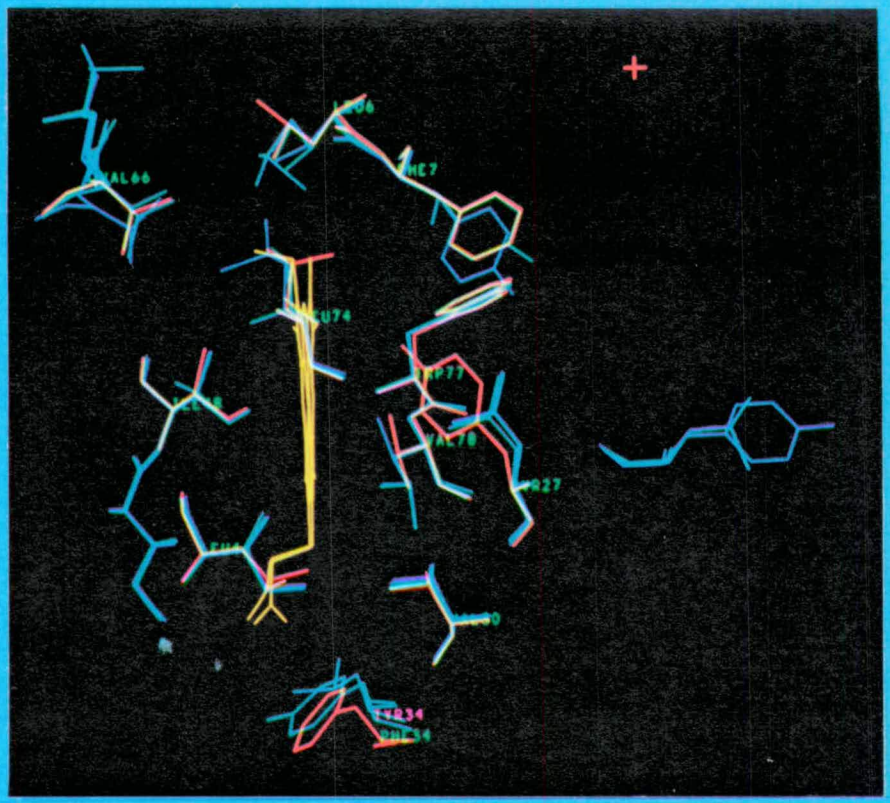


Figure 6.15.b

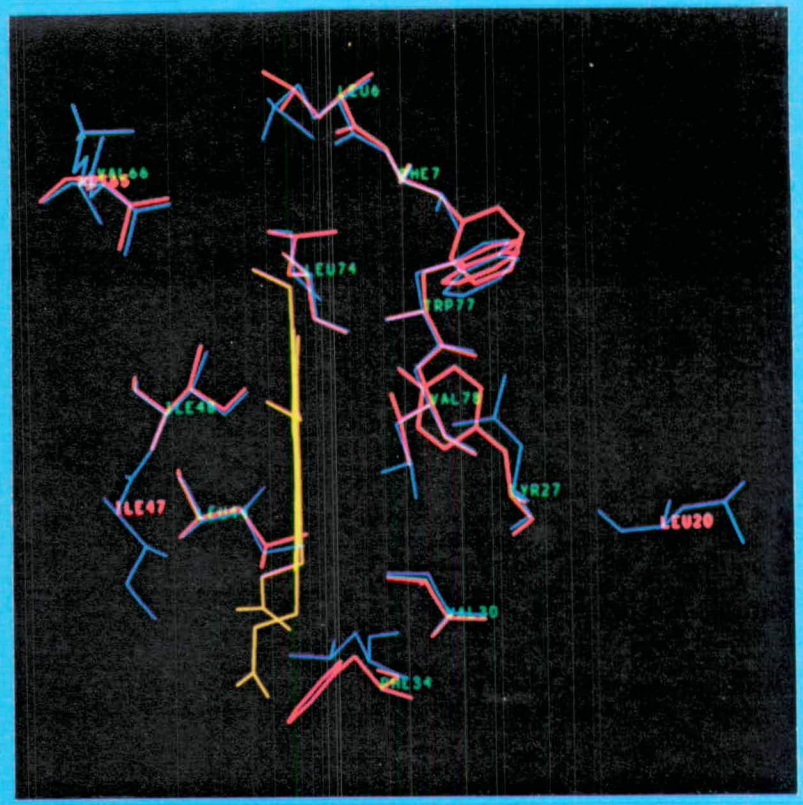
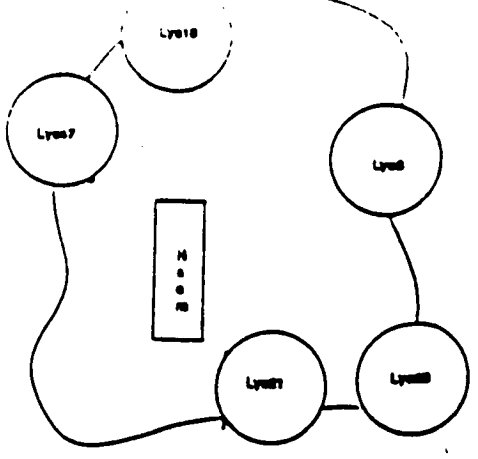


Figure 6.15. Hydrophobic amino acids in *Pseudomonas aeruginosa* cytochrome c551 and *Bacillus* cytochromes c superposed.

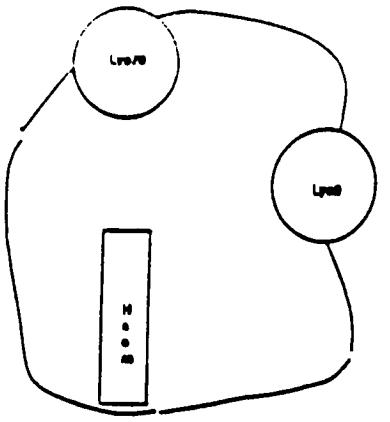
The structures are viewed from the front with the haem propionates pointing downwards.
Numbers refer to positions in the *Pseudomonas* structure (figures 6.10 and 6.12). Hydrophobic amino acids in the left side loop of the *Pseudomonas aeruginosa* structure, which closes off the haem crevice, are not shown.

Figure 6.15a: ***Pseudomonas aeruginosa*** **c551,** ***Bacillus azotoformans*** **P2-c552 and** ***Bacillus. licheniformis*** **c552.**
Red side chains are in the *P. aeruginosa* cytochrome c551. Blue side chaine are in the *Bacillus azotoformans* cytochrome P2-c552.
Bluegreen sidechains are in the *Bacillus licheniformis* cytochrome c552.

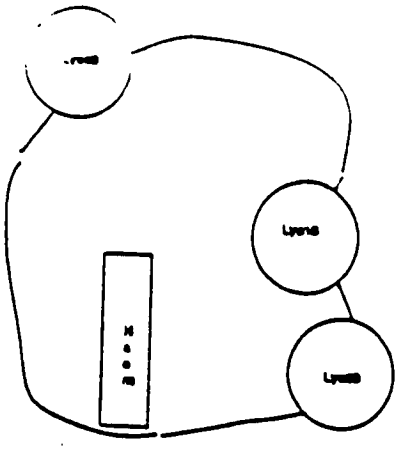
Figure 6.15b: ***Pseudomonas aeruginosa*** **c551 and** ***Bacillus azotoformans*** **P2-c552.**
Red side chains are in the *P. aeruginosa* cytochrome c551. Blue side chains are in the *Bacillus azotoformans* cytochrome P2-c552.



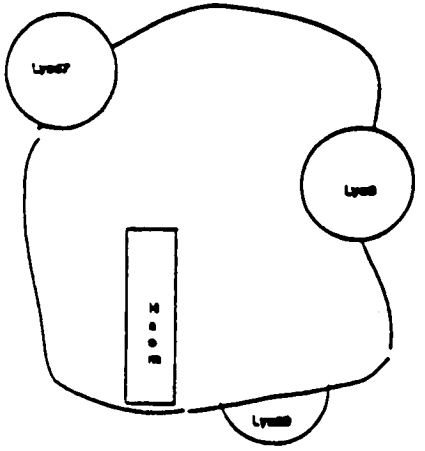
A. *Pseudomonas aeruginosa*



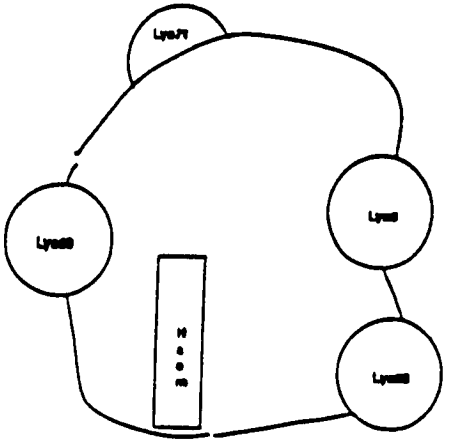
B. *Bacillus azotoformans*



C. *Bacillus licheniformis*



D. *Bacillus* PS3



E. *Bacillus subtilis*

Figure 6.16. Schematic diagrams of the distributrion of basic residues on the surface of the *Ps. aeruginosa* and four homologous Bacl sequences from the genus *Bacillus*.

The cytochromes are viewed from the front, onto the edge of the haem with the propionates pointing downwards. The diagram of the *Ps. aeruginosa* c551 structure is adapted from Moore and Pettigrew (1990).
Residue numbers refer to the *Ps. aeruginosa* c551 structure (figures 6.12 and 6.10).

6.5. THE HERMENEUTICS OF CYTOCHROMES C OF CLASS I

6.5.1. Protein Taxonomy

'Biologists should realise that before long we shall have a subject which might be called "protein taxonomy" - the study of the amino acid sequences of the proteins of an organism, and the comparison of them between species" (Crick, 1958)."Sequence space is not uniformly occupied" (Ambler, 1991)

Protein taxonomy has mainly been approached from the vantage point of organism (bacterial) phylogeny. However, with increasing numbers of protein sequences available, emphasis in protein analysis has shifted away from evolutionary and phylogenetic considerations to structural and functional ones. Such analysis can be termed taxonomic. It has a broader scope than phylogenetic construction based only on cladistic analysis of homologous sequences.

Protein taxonomy can be defined in similar terms as Sneath (1989) does for bacterial taxonomy. The primary aim of protein taxonomy is to give a classification that is useful for a variety of purposes, including identification and to produce data-bases that summarise as much relevant information about them as possible. In this context, by similar paraphrasing of Sneath: Classification would be the ordering of proteins into groups on the basis of their relationships, not necessarily confined to ancestry.

Taxonomic analysis of proteins is not restricted to homologous proteins, where there is a relatively smooth transition from one sequence to another. In a sense, its subject is a matter of definition in each case. It could, in theory, be all proteins or all cytochromes or only cytochromes c of Class I. Perhaps a too broad generalisation is of limited value, but in a sense the entire field of protein sequence and structure analysis can be put into the domain protein taxonomy.

Obviously, taxonomic analysis of protein sequences also encompasses phylogenetic or evolutionary reconstruction. Besides a possible cladistic ordering of homologous proteins, rational proposals or hypotheses can be made on the evolution of protein structure or the metabolic pathways in which the proteins play a part. The construction of protein family trees with historical and evolutionary dimension is, however, only one of the long time goals.

Taxonomic grouping of sequences should obviously be based on sequence characteristics. The sequence of steps in such an analysis procedure has been defined by Sneath (1987). A slightly altered version is given below.

-Firstly, a similarity must be recognised, which requires alignment of the sequences to reveal these matching regions. Sequence classes may be delimited on the basis of common characteristics that are defined by the group and which exclude other sequences. In other words, sameness is established.

-Secondly, when possible and sensible, sequence similarity is expressed numerically as distances or similarity indexes, based on the proportion of matches and mismatches between the sequences.

-Thirdly, the relationships are summarised by some method to reveal its taxonomic structure and its implications. Whether the results are phenetic or cladistic depends on the kind of sequence data and the methods used.

At the first level in taxonomic analysis of proteins, described above, sequential relationships are defined solely on the basis of the similarities of primary structures, where distances are derived by some numerical method, from pairwise comparisons. At a higher level, they might also include all other information that is coded, and quantifiable in the sequence, whether it be certain sequence motifs or secondary and tertiary structural characteristics.

6.5.2. Classification of Bacterial Cytochromes

Cytochromes c are distinguished from other cytochromes by the covalent attachment of the haem. It has a characteristic absorption spectrum by which they can be detected and further grouped into distinct structurally unrelated classes. However, classification based on spectra as well as on other biophysical characteristics or function is clearly secondary to one based on primary structure if evolutionary or phylogenetic relationships and structural affinities are to be evaluated.

With few exceptions the more than one hundred c-type cytochromes sequences now known fall into at least three structurally unrelated classes that were recognised by Ambler ((1977), table 6.2). They all have characteristic haem attachment sequence, C–CH, but apart from that they are very dissimilar. ClassI is further divided into subclasses with distinct sequence characteristics.

TABLE 6.2.Classification schemes of cytochromes

A: Classification scheme of Ambler (1977).
P: The approved nomenclature adopted and revised by Pettigrew and Moore (1987).
D: Classification scheme of Dickerson (1980).

Information on the main characteristics of the cytochromes c are compiled from Ambler (1982), and Pettigrew and Moore (1987).

A	P	D

CLASS I

Low spin, His + Met haem coordination, haem near the N-terminus, 80 -120 amino acids. This class is subdivided into:

a) CLASS I LARGE.

Loop of residues closes off the bottom of haem crevice

Subclass IA. C2 L

Long cytochrome *c2*
97 -124 aa.Size variation at surface loops.
Spectra: α(550 nm) symmetrical α/β = 1.6.
E_m= 290-390 mV. 30% minimum pairwise sequence identity. Contain several extra loops compared to class IB cytochromes, although three dimensional structures are generally similar. The insertions and deletions are in different places in the different members of this class.
Examples: *Rhodospirillum rubrum* cytochrome c2, *Aquaspirillum itersonii* cytochrome c-550.

Subclass IB C2 M

Mitochondrial cytochrome *c*.
103-112 aa. Size variation at N-terminus.
Spectra: α(550 nm) symmetrical. α/β= 1.8
E_m= 244-264 mV.40% minimum identity.
Examples: *Rhodospeudomonas globiformis* cytochrome c2, *Nitrobacter agilis*.

b) CLASS I SMALL.

Lacks loop of (a) Left side folds downwards to close off the haem crevice.

Subclass IC C4

a) c4-cytochromes.
One or two domain structure with one haem per domain.80-90 aa. per domain. Spectra: a split or widened α-band, 550-554 nm. α/β= 1.2 .
Low 260-280 absorption, no shoulder at 290 nm.
E_m= 190-320 mV.40% minimum identity. No tryptophan. Sequence characteristics, at the sixth ligand -M------LS---I-----Y- and aromatic and proline residues after the haem binding site. Lacks a loop at the base of the haem.
Examples: *Thiobacillus neapolitanus* c554(547). pseudodimeric cytochrome c554(548) from a halophilic *Paracoccus* and *Azotobacter vinelandii* dihaem cytochrome c4.

b) algal cytochromes (soluble f). C6 S*

83-89 aa. Spectra: α asymmetric (552-554).Em=335-390
38% minimum identity. Lacks loop at base of haem.

Subclass ID

Cytochrome c_8.
70-83 aa. Spectra: a (551-553 nm) symmetrical or asymmetrical, a/b= 1.7-2.1 . E_m: The majority falls between 170-270 mM, however, *Rhodocyclus tenue* cytochrome c553 has E_m= 405 mM and *Rhodocyclus gelatinosa* E_m= 20 mM. 35% minimum identity. A conserved tryptophan at the c terminal and another one, four amino acids before the sixth ligand methionine. Several proline residues around the sixth ligand All, except perhaps the *Rhodocyclus* cytochromes lack a loop at the base of the haem.Examples: *Pseudomonas* cytochrome c551, *Rhodocyclus tenue* cytochrome c553, *Hydrogenobacter thermophilus* cytochrome c552.

Class IE C5

Cytochrome c_5.
78-116 aa. Size variation at the N-terminus. Spectra: α(554 nm) symmetrical α/β = 1.4. E_m= 265-320 mV. 39% minimum pairwise sequence identity. Contains a sequence C--C at the C-terminus which forms a short dishulphide loop Lacks a loop at the base of the haem. A ragged N-terminus.
Examples: *Pseudomonas mendocina* cytochrome c5 organism H-1-R cytochrome c5

Unclassified.

Cytochrome c from green sulfur bacteria
Cytochrmoe c555 (*Chlorobiocea*).
86-99 aa.Size variation at N-terminal. α:(555 nm) asymmetric. α/β = 1.2. 54% sequence identity. Lacks loop at the base of haem. The pair of cysteines observed in the c5 sequences is missing.
Example *Chlorobium limicola* c555.

UNCLASSIFIED: C1

Haem near N-terminus. A putative class I cytochrome c.Cytochrome c1 from the alpha division of purple bacteria and mitochondria, and cytochrome c f6 from cyanogen bacteria and chloroplasts. Membrane bound.These cytochromes form a part of the quinol reductase complex,with a cytochrome b and the Rieske protein. Size: 32 kD-6kmD. Redox potential, c1: approx 250 f: approx 400.

CLASS II C8

Haem near C terminus.

a) high spin.

Histidine only haem coordination.
Example: The cytochromes c'

b) low spin.

Histidine + methionine coordination of the haem.
Examples: *Rhodospuedomonas palustris* cytochr. c556.

CLASS III C3

Multihaem, one per 30 -40 amino acids, bis histidyl haem coordination. Further subdivsions are based on the numbers of haem.
cyt c_3 (4 haem)
cyt cc_3 (8 haem)
cytochrome c_7 (3 haem)

The diversity of cytochromes c is manifest both within the same species where different cytochromes c classes and subclasses coexist, as well as between species and higher taxonomic levels (Ambler, 1977, 1991).

6.5.3. Classification of Class I Cytochromes

The great majority of cytochromes c, about two thirds, belong to Class I, most of which have been grouped into at least five subclasses defined by Ambler in 1977 and further revised in 1991 (table 6.2). The classification is based on size, characteristic insertions or deletions and invariant or near invariant amino acids. A minimum is around 35 % identity within class, but when subclass lines are crossed, identity falls to a level expected to be found by chance. A more general classification scheme of Class I cytochromes, was presented by Dickerson (1980) where besides sequence similarity he places emphasis also on size and three dimensional characteristics.

The evolutionary significance or structural attributes of a new cytochrome c sequence have to be evaluated in the context of previously accumulated sequence information on cytochromes c. Many of the recently sequenced cytochromes, including those from the genus of *Bacillus*, have not been classified yet. A proposal for an overall alignment of cytochromes c is made in figure 6.17a. This alignment includes the hitherto unclassified cytochromes c and their affinities to already defined cytochrome groups or subclasses are evaluated.

The alignment in figure 6.17a includes representatives of all the bacterial subclasses (see table 6.3). It also includes sequences from two eukaryotic organisms, the chloroplast of the red algae, *Porphyra tenera* and tuna mitochondrial cytochrome c. When possible, two or more highly divergent sequences of the same subclass are presented for visual enhancement of its sequence characteristics and to reduce the effect of random noise in the subsequent UPGMA clustering. It helps to distinguish between significant and superficial similarities between the subclass sequences. The alignment includes most of the published sequences that have not been assigned a group or taxonomically evaluated to any great extent. These are: *Bacillus* PS3 c(aa3), two BacI sequences from *Bacillus azotoformans* and *Bacillus* PS3, the *Chlorobium limicola* c555, the *Pseudomonas putida* c553, the *Desulfovibrio vulgaris* c553, the *Thermus thermophilus* c552, the *Methylococcus capsulata* c555, the *Pseudomonas*

cd1 sequences, the *Methylobacter extorquens* CL and the *Pseudomonas aeruginosa* cytochrome CPO.

The alignment was based on:

1) Previous alignments within subclasses, mainly by Ambler, for defining positional conservation of physicochemical properties.

2) The crystallographic data of the cytochrome c5 from *Azotobacter vinelandii*, cytochrome c551 from *Ps. aeruginosa* the c2 cytochrome c from tuna, The cytochrome c553 from *Desulfovibrio vulgaris*, the c553 cytochrome from *Anacystis nidulans* and c555 from *Chlorobium limicola* (previously *Chl. thiosulphatophilum*)

The following procedure was used:

1) The structures were anchored at those positions which are generally considered invariant or highly conserved in cytochromes c of Class I. These are: Glycine in position three, aromatic amino acid in position seven, the haem site C--CH in positions 12-16, proline in position 33 (however, see below for the c4 sequences), the methionine in position 87, the acidic region 98-100 and the aromatic acid in position 108.

2) The secondary structure core elements were delimited on the basis of available crystallographic data. For those sequences with no available three dimensional structure, the core elements were delimited by glycine and aspargine rich regions. These regions were considered indicative of turns or the start of loops or coil regions.

3) Further alignment of secondary structure elements aimed at optimising the adherence to the generally conserved hydrophobicity pattern of cytochromes c, with a special emphasis on the conservation of bulky aromatic amino acids.

4) Indels were kept to a minimum in core elements, and also in the loops. Indels were only introduced in a sequence, if it resulted in marked increase in identity to a number of apparently important structural positions of other sequences.

The rationale behind this particular alignment and the difference to other proposals is discussed below in sections 6.5.3.1.3 and 6.5.3.2. The subsequent grouping of cytochromes based on this alignment is discussed in sections 6.5.4.1, 6.5.4.2 and 6.5.4.3.

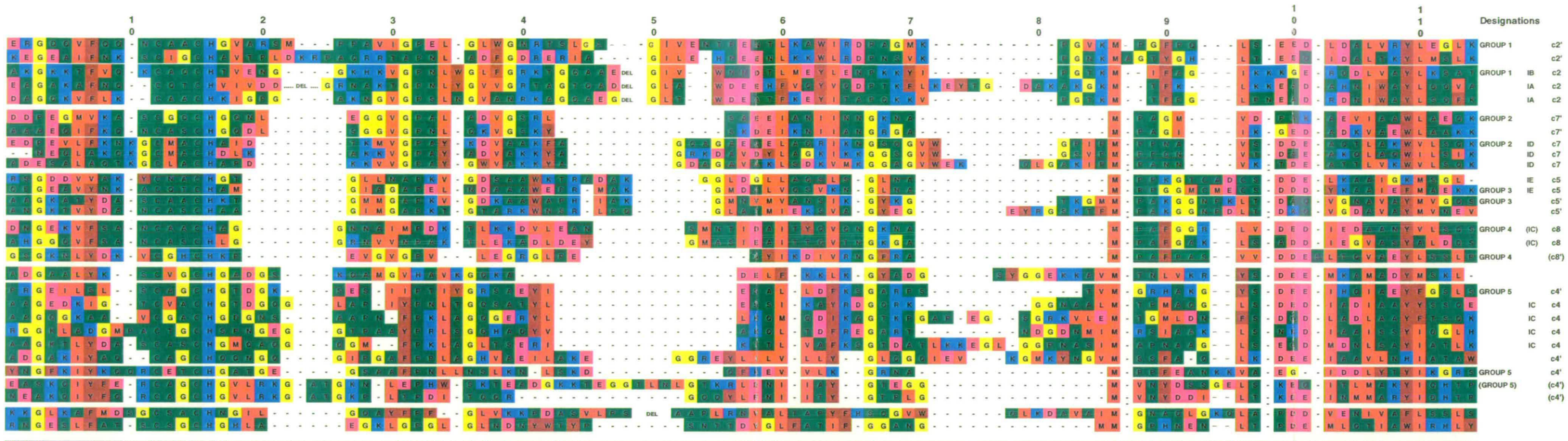


Figure 6.17a. An overall alignment of cytochrome c sequences of class I.

The sequence numbers refer to table 6.4 below.
Only the core elements of the cytochromes are shown, but not extended N- and C-terminal sequence parts . Del means that amino acids in this particular region are not shown. Designations refer to table 6.4 below.

Hydrophobic aminoa acids | Aromatic amino acids | Basic amino acids | Acidic amino acids | Ambivalent amino acids | Glycine

Table 6.4. Taxonomic ordering of cytochrome c sequences of class I.

Nr	Species	Cyt.	Group designation		Nomenklature	Group	Nomenklature
			Ambler	Dickerson	Moore & Pettigrew	This work	This wor
					adopted & proposed	Proposed	Proposed
1	*Thermus thermophilus*	c(aa3)				1	c2'
2	*Bacillus* PS3	c(aa3)				1	c2'
3	Tuna cytochrome c	c2	IB	M	c2	1	
4	*Rhodobacter spaeroides*	c2	IA	L	c2	1	
5	*Rhodopseudomonas acidophila*	c2	IA	M	c2	1	
6	*Bacillus azotoformans*	c552				2	c7'
7	*Bacillus* PS3	c550				2	c7'
8	*Pseudomonas aeruginosa*	c551	ID	S	c7	2	
9	*Hydrogenobacter thermophilus*	c552	ID	S	c7	2	
10	*Rhodocyklus purpureus*	c553	ID	S	c7	2	
11	*Azotobacter vinelandii*	c5	IE		c5	3	
12	HIR	c5	IE		c5	3	
13	*Chlorobium limicola*	c555				3	c5'
14	*Prostecochloris aestuarii*	c555				3	c5'
15	*Porphyra tenera*	c553	IC	S*	c6	4	

Nr	Species	Cyt	Subclass		Nomenclature	Group	Nomenklature
			Ambler	Dickerson	Moore & Pettigrew	This work	This work
					adopted & proposed	Proposed	Proposed
16	*Anacystis nidulans*	c553	IC	S*	c6	4	
17	*Pseudomonas putida*	FC				(4)	
18	*Desulfovibrio vulgaris*	c553					
19	*Chlorobium limicola*	FC				5	c4'
20	*Paracoccus denitrificans*	c4	IC		c4	5	
21	*Azotobacter vinelandii*	c4	IC		c4	5	
22	*Azotobacter vinelandii*	c4	IC		c4	5	
23	*Thibacillus neapolitans*	c4	IC		c4	5	
24	*Thermus thermophilus*	c552				5	c4'
25	*Methylcoccus capsulata*	c555				5	c4'
26	*Pseudomonas stutzeri*	cd1				(5)	
27	*Pseudomonas aeruginosa*	cd1				(5)	
28	*Methylobacter extorquens*	CL					
29	*Pseudomonas aeruginosa*	CCP					

IA -IE: Subclass designations of Ambler 1977 and 1991. L= large, M=medium, S=small, S*=small: Class designations of Dickerson 1980. c2,c4,c5,c7, c8: The approved nomenklature adopted and proposed by Moore and Pettigrew and Moore,1990.

1: Mather et al., 1991; 2: Ishizuka et al., 1990; 3: Kreil 1963, 1965, (cited in Moore and Pettigrew 1990); 4: Ambler et al., 1979; 5: Ambler et al., 1979; 6: This work; 7: Fujiwara et al., 1993; 8: Ambler, 1963; 9: Sanbogi et al., 1989; 10: Ambler, 1991; 11: Carter et al., 1985 (cited in Moore and Pettigrew, 1990); 12: Ambler, 1991; 13: van Beeumen et al., 1976 (cited in Moore and Pettigrew, 1990); 14: van Beeumen et al., 1976 (cited in Moore and Pettigrew, 1990); 15: Ambler and Bartsch, 1975 (cited in Moore and Pettigrew, 1990) 16: Margoliash (cited in Dickerson, 1980); 17: (MacIntire et al., 1986 (cited in Moore and Pettigrew, 1990); 18: Nakan et al., 1983 (cited in Moore and Pettigrew, 1990); 19: van Beeumen et al., 1991; 20: Ambler et al., 1978; 21: Ambler et al., 1984; 22: Ambler et al., 1984; 23: Ambler et al, 1985; 24: Titani et al., 1985; 25: Ambler et al., 1986; 26: Smith et al., 1991; 27 :Smith et al., 1991;; 28: Nunn et al., 1988 (cited in Moore and Pettigrew, 1990) 29: Ronnberg et al., 1987 (cited in Moore and Pettigrew, 1990).

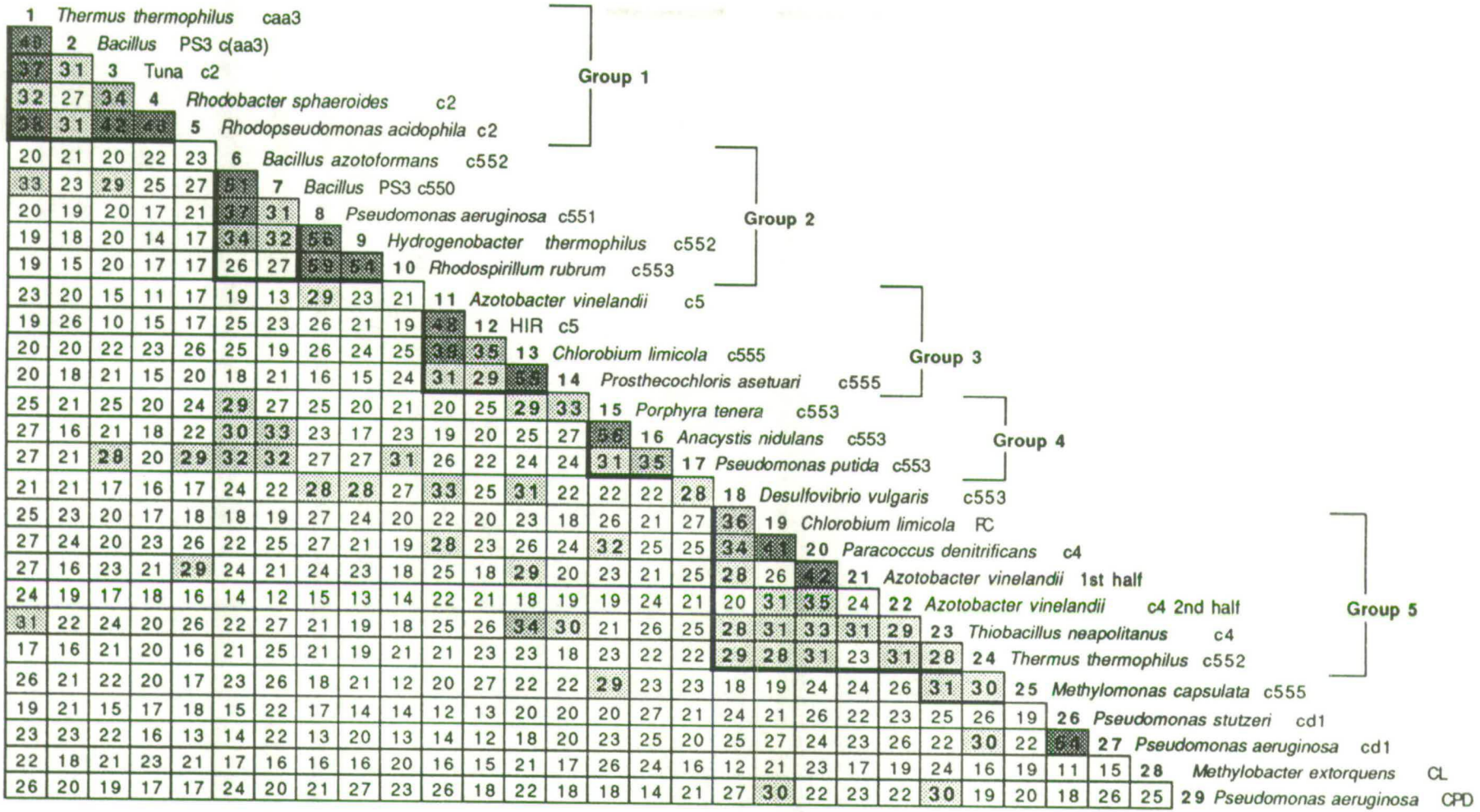


Table 6.17b. Matrix of % identities for the overall alignment of cytochromes c in figure 6.17a.

The values are percentage identities of minimum common sequence length.

28% - 30%
31% - 33%
34% - 36%
> 36%

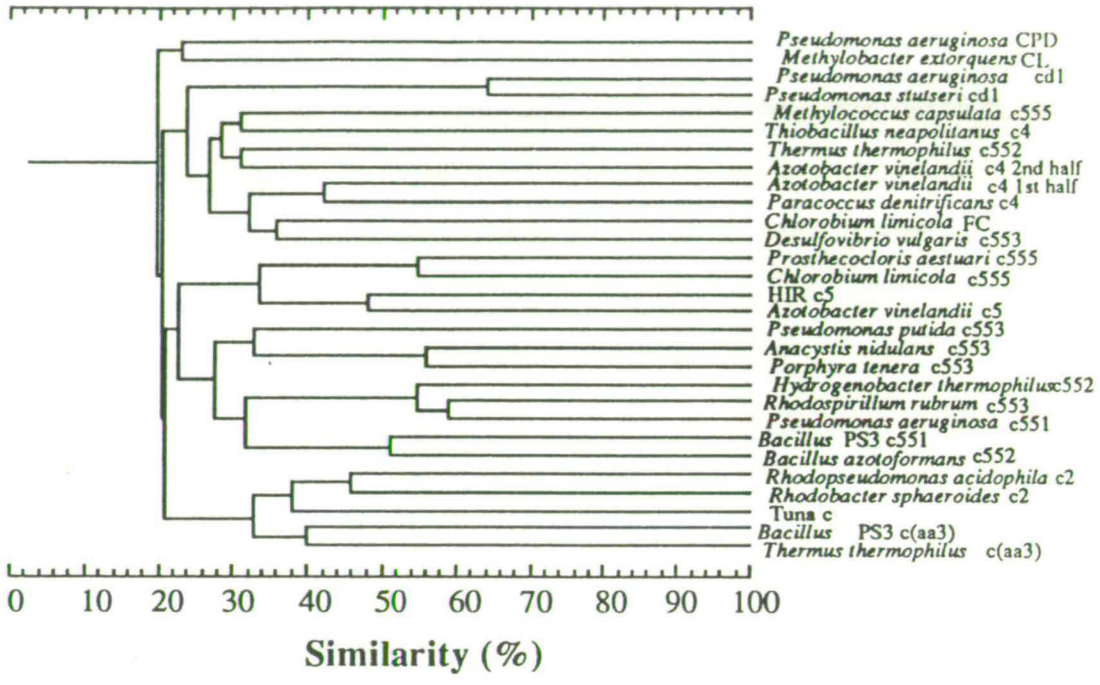


Figure. 6.17c. An UPGMA (%) similarity dendrogram cast from the identity matrix in figure 16.7b, of the overall alignment of cytochome c sequences in figure 6.17a.

6.5.3.1. Alignment of distantly related protein sequences

The prerequisite for meaningful comparative analysis of a group of sequences is a correct alignment. The proper alignment of primary structures of proteins is not necessarily the mathematically optimal one on the bases of identity or weighted score matrices. The whole informational content of the sequences has to be taken into account and the alignment must reflect three dimensional features of the proteins or, according to a good taxonomic practice, sameness has to be established before the definite similarity or distance value is numerically calculated. This will not necessarily lead to the correct historical alignment that is required for phylogenetic reconstruction, but then it is also beyond reach.

6.5.3.1.1. Alignment of distantly related primary sequences of cytochromes c by numerical methods

Numerical methods of comparing primary structures of proteins based exclusively on sequence similarity, especially if they are distantly related, may not have any biological significance. The maximum identity or calculated score obtainable is very often at the level which can be obtained by chance (25% identity for proteins). Also the building blocks of the cytochromes or secondary structure elements show a predominant periodicity of amino acids of specific physicochemical properties. These patterns may, in some cases, be picked out in similarity analyses and thus regions with different position or roles in the overall three dimensional structure may be aligned. When comparing distantly related cytochromes, historically unwarranted gaps are often introduced into the core element regions to maximise identity. Furthermore, while indel events are more likely to occur in loop regions, these regions may not be homologous in different subclasses of cytochromes c, as their presence in different sequences may represent separate evolutionary events. Their alignment, maximised in identity and perhaps structurally valid because of convergence, will skew all historical interpretation.

6.5.3.1.2. Alignments of distantly related cytochromes c based on three dimensional structures

Alignment on the basis of three dimensional superposition of cytochrome c structures is perhaps more reliable than alignment based on sequence similarity. The credibility of such an alignment, however, depends on the quality of the data. Differences have been observed

between authors in defining secondary structures from crystallographic data as in bovine pancreatic trypsin inhibitor, in which 12 residues of 58 were differently assigned by Robson and Suzuki, and Levitt and Greer (Garnier and Robson, 1989). A similar case for cytochromes might be the different spatial assignment of the residues in *Rhodospirillum rubrum* cytochrome c2 by Salemme *et al.*, (1973) and Bathia *et al.*,(1984), respectively. The sequence of the one resolved by Salemme does not pack around the haem as would be expected on the basis of requirements for the conservation of the haem pocket. (Chothia and Lesk, 1984). Such discrepancies appear because of the different refinements of the structures and because of different criteria for the definition of secondary structures (Garnier and Robson, 1989) .

The very different structural features of the *Chl. limicola* c555 and *Azotobacter vinelandii* c5 cytochromes (figures 6.18e,f) despite extensive sequence similarity, may be explained on similar grounds. The only major difference is the presence of a larger left side loop preceding the methionine site. The new c5 sequence from H1R further (figure 6.17a) supports this. The hydrophobicity pattern is almost identical with bulky, mostly identical, aromatic acids in the same generally conserved positions. No indel events have to be proposed for optimal alignment except in the above mentioned left side loop region. Despite this, the sequence part from 38 to 48 (figure 6.17a), involving a conserved tryptophan in both structures, is defined in helix formation on the right side of the haem in the c5 structure (Carter *et al.*, 1985), but as a coil region in the c555 structure, which passes behind the edge of the haem (Korszun and Salemme, 1977; Salemme, 1977). Furthermore, the loop region preceding the proline, which hydrogen bonds with the fifth ligand histidine, has a different conformation in the two structures. It turns upwards in c5, but downwards in the c555 structure. The c5 conformation is found in most other cytochrome c crystallographic structures in one form or another(6.18a,b,c,d,e,f; the bow region is blackened or designated B in these figures). The *Chlorobium limicola* c555 crystal structure had resolution of only 3 Angstrom, which may explain these discrepancies.

In figure 6.19 are different alignments based on crystal structures of cytochromes c. It is interesting to compare the alignments of Nakagawa *et al.* (1990) and Carter *et al.* (1985) the sequences of *Pseudomonas aeruginosa* c551 and *Azotobacter vinelandii* c5 based on the same crystallographic data. They do not agree in many aspects (figure 6.19). Amino acid positional equivalency in alignments is based on superposition of α-

carbons of three dimensional crystallographic structures, but amino acids that align spatially need not be homologous. It is often apparent, with the derived alignments of the corresponding primary structures, that gaps have had to be introduced into the primary structure to observe the three dimensional spatial equivalency of residues. These gaps are often indiscriminately placed, with little regard to the observation that core elements appear to be "evolutionarily more stable structures" than loop regions (Gribskov *et al.*, 1990). Also, there are examples of alignments, where regular secondary structures are aligned with loop regions.

The criteria of alignment of sequences by superposition, leads to a similarity index based on spatial equivalency. It may not be the best historical alignment and perhaps not the best structural alignment either. It may even be wrong.

6.5.3.1.3. Alignment based on conserved amino acids and/or sequence motifs

It has been observed, that distantly related proteins have similar characteristic patterns of conserved residues (Patthy, 1987; Gribskov *et al.*, 1987). The residues are either identical or they have similar physicochemical properties. The sequence patterns can be defined on the basis of multiple alignments of recognisably homologous sequences. They reveal that regions or even single sequence positions in the primary structures tolerate variation to a different degree. Also, there are regions which tolerate gaps or insertions to a different extent. In the following discussion a distinction will be made between sequence motifs and signatures. **A sequence motif is defined as a characteristic amino acid pattern of a small 'structural unit' of a primary structure. A signature is defined as the distinctive conserved amino acid pattern of the entire set of sequences which comprise the group.** Thus a signature of a sequence group comprises the cumulated sequence motifs recognised in that group and which are revealed by the alignment. It follows, that some, but not all sequence motifs can be shared between two groups with different signatures. When two distinct, but related, groups of sequences are merged, the signature characterising the new enlarged group will be composed of the shared sequence motifs.

DNA sequence motifs have been used in the analyses of 16S rRNA, for evaluating phylogenetic affinities (Woese, 1987). Distinct sequence motifs in proteins are also used more or less intuitively in aligning one sequence to another. Sequence motifs can represent either homology or may have

risen in unrelated proteins by convergence. Sequence motifs of the latter kind, may define active sites in enzymes, as in the structurally unrelated serine proteases. It can also define a specific structural property of the protein, such as the characteristic C--CH sequence to which a haem is bound and which thereby defines a cytochrome c. This haem binding sequence motif is found in cytochrome c sequences of all the apparently unrelated cytochrome c classes.

Chothia and Lesk (1984, 1986, 1987) have analysed the core structures of different protein families including cytochromes c. They observed that secondary structure elements are shifted relative to each other in different proteins of the same family and especially in alpha-helical proteins. Thus the root mean square differences in positions of topologically equivalent residues, in a pair of homologous proteins, increases with percentage difference in sequence. Positions of buried amino acids differ less in homologous structures (Blundell *et al.*, 1988) and similar spatial positions are obtained by coupled mutations in the secondary structure elements. This conserved pattern of hydrophobic residues may provide a general framework or a blueprint, on which to base alignments in protein regions without strong sequence motifs, provided that the sequence can be anchored by residues in adjacent regions of more distinct sequence motifs.

It is reasonable to assume that there are only limited numbers of ways to efficiently pack the helix (or core elements) blocks of cytochromes c around the haem. The packing must predominantly be constrained by the requirements for the conservation of the haem pocket, which may require that hydrophobic residues are placed in more or less fixed positions relative to the haem. The smaller the polypeptide chain is, the fewer possibilities are for attaining an efficient packing. Those sequences may therefore be expected to converge upon similar pattern of conserved hydrophobic amino acids.

A signature can be defined for cytochromes c of Class I. Many of these sequence characteristics have been pointed out by Ambler (1991), and Moore and Pettigrew (1990). The most notable are the following (figure 6.17a): Glycine in position 3, hydrophobic-aromatic residues in positions 6 and 7, The haem site sequence C--CH in positions 12 15 and 16, the proline in position 32, which hydrogen bonds to the fifth ligand histidine, the sixth ligand methionine in position 86 the triplet of acidic residues in positions 97,98 and 99 and the adjacent hydrophobic and aromatic residues in positions 107 and 108.

It may be possible to add the following sequence motifs to the general signature sequence of cytochromes c: firstly, the alpha helical motif of the third helix, the hydrophobic residues in positions 60 and 64 and the glycine in position 67. Secondly, the proline which hydrogen bonds to the fifth ligand histidine can perhaps be included in the more extensive sequence motif, from position 32 to 37, of -**Pro-X-(h.ph)-(X)-Gly-** (**X** stands for any amino acid, **(X)** means one amino acid or an indel event and **(h.ph)** stands for hydrophobic amino acid). The proline is placed upwards into the right distance and in configuration for hydrogen bonding to the fifth ligand histidine. The hydrophobic residue, **(h.ph)**, may form the upper right corner of the haem pocket and the glycine may act as a breaker from the proline site to send the backbone chain away downwards to connect to the next core element (the second helix region in the *Ps. aeruginosa* c551). The major exception to the sequence motif of the proline site, is in the *Anacystis nidulans* sequence. Instead of the hydrophobic residue there is a lysine in position 34. Also, the proline residue in position 32, is not conserved in the cyanobacterial cytochromes, so that the anchoring obtained by the histidine-proline hydrogen bonding is not made. This is supported by crystallographic analysis of the *Anacystis nidulans* structure (Ludwig *et al.*, 1982). Instead of turning upwards towards the histidine, this proline goes downwards (figure 6.18). Some sequence motifs are very well conserved in residue identity within the cytochrome c Class I sequence family. Some of these motifs define subclasses (Ambler 1977, 1982; 1991, table 6.2) to the exclusion of other sequences. Thus the IE subclass cytochromes c (c5) have besides other common features, a characteristic di-cysteine sequence C--C closely following the sixth ligand methionine (figure 6.17a). It is also noteworthy that the best conserved sequence motifs within a subclass, very often follow closely the general conserved pattern of hydrophobic residues for the whole family. This is in accordance with the observation of Chothia, mentioned above, that buried hydrophobic amino acids are generally much better conserved than residues of other types. Ambler (1982, 1991), has for example defined a sequence motif for the C-terminal part of the IC subclass cytochromes as -M------LS---I-----Y-. This sequence motive is conserved in identity in the majority of the sequences of this specific subclass and most sequences outside it have distinctly different residues, although predominantly hydrophobic ones, in the same positions. However, sequences have recently been reported, that cannot be confidently placed into the IC subclass and which have the same or very

similar C-terminal sequence motif. An example of such a sequence is the recently sequenced C(aa3) cytochrome from *Bacillus* PS3, which has the pattern -M------LT---I-----Y- and is otherwise very similar to the c2 cytochromes. At a further end of the scale some sequence motifs are so general that they cannot be defined as specific to any particular protein group, unless in the context of the overall sequence information. These motifs may include, for example, alpha helix sequence patterns, which are common or similar for secondary structural elements of proteins of entirely different nature and overall structure.

The IC subclass C-terminal pattern and the PS c(aa3) cytochrome may serve to illustrate the difference emphasised between sequence motif and a signature. The IC subclass includes both the algal soluble f sequences and the so called c4 cytochromes. They share the above C terminal sequence motif, but apart from this sequence motif they are very dissimilar and can be said to have distinctly dissimilar signatures.

Whether sequence motifs or signatures are defined by residue identity or positional conservation of physicochemical character of the amino acids, they are rarely, or perhaps never, absolutely conserved. Deviations are usually revealed when a number of sequences of a specific group is examined. Notable deviations have even been found to the highly conserved C--CH haem binding sequence of cytochromes c. Thus the first cysteine is missing in the *Euglena gracilis* cytochrome c and cytochromes with an additional residue between the cysteines, C---CH, have been found (van Beeumen, 1991). The questions arise as to how much deviation is acceptable, and can these sequence motifs be expressed numerically and quantified to eliminate the subjectivity in defining them. Perhaps profile analysis can be used. In figure 6.8 the positional conservation is plotted for the BacI and the ID subclasses and for guidance the standard deviation from the mean positional conservation for the entire proteins. It is obvious from the graph that certain positions are highly conserved while others are not. The definition of sequence motif for a sequence or sequence part can thus be defined in relation to the standard deviation from the mean positional score for the whole protein group. However, this is only possible if a large number of sequences are available for a particular protein structure.

The credibility of using sequence motifs or signatures for alignment is supported by the results obtained by various authors, who have devised methods, based on patterns of positionally conserved amino acids, for searching protein databases for distant but biologically significant

relationships. The signal to noise of such searches was significantly reduced and permitted sensitive detection of distantly related members of a homology group. These sequences similarities had not been detected by more traditional methods, based on similarity scoring to a single sequence (Patthy, 1987; Gribskov *et al.*, 1987).

6.5.3.2. Different alignments

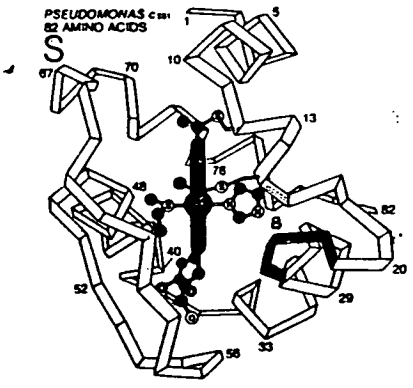
Figure 6.20 shows the alignment of Moore and Pettigrew (1990), van Beeumen *et al.* (1991), a selection of sequences in the alignment of Dickerson (1980) and the alignment presented in this thesis (figure 6.17a) for comparison. The same sequences in the different alignment have the same colour to facilitate the comparisons. The alignment presented in this thesis differs from these alignments in many aspects. The major difference to the Pettigrew and Moore alignment is in the third helix region. This region is occasionally aligned with loop regions of other sequences. The arguments for the alignment presented in this thesis is given in the preceding chapters. In addition, Chothia and Lesk (1984, 1986, 1987) have concluded from sequence and three dimensional structure comparisons of sequences from different protein families, that major shifts can apparently occur in the placement of alpha-helices due to mutations in the intervening regions. This has to be guarded against, as non homologous regions may therefore occasionally occupy the same space or 'align' spatially when the three dimensional structures are superposed. While neither alignment can be proven to be the right historical alignment, it is perhaps more logical to expect that core regions tend to evolve into core regions rather into loops and converse. According to this a closer relationship is proposed for the c5 and c555 sequences, which are then aligned to the ID sequence according to the proposed alignment apparently based on crystallographic evidence.

The alignment of the conserved prolines in position 32 is not the one favoured by Moore and Pettigrew for the *Thermus thermophilus* c-552 and *Ps. aeruginosa* cytochrome c peroxidase, and Dickerson for the c4 cytochromes. Instead a proline closer to the haem is put into this position. This proline is, however, not conserved in all sequences of the c4 cytochromes and, furthermore, it is not found in one of two cd1 cytochromes c that appears to be structurally related to the IC sequences and especially to the c4 cytochromes. Furthermore, the second proline region conforms to the proline site motive, as defined above, -pro-(x)- (x)-

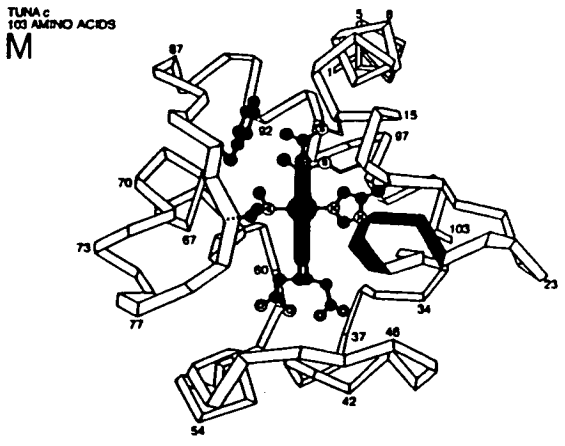
(hydrophobic amino acid)-gly-. It is proposed that this sequence motif corresponds to the upper right hand corner side of the haem pocket.

The *Desulfovibrio* alignment is based on the recently published crystallographic structure (Nakagawa *et al.*, 1990) and the derived alignment, differs in some major aspects from the Pettigrew and Moore, and Dickerson's alignments. Thus, the histidine in position 32, appears to be spatially equivalent to the conserved proline, which in other cytochromes of Class I, hydrogen bonds to the fifth ligand histidine. Also, the first loop appears to take the place and role of the second helix region in other structures.

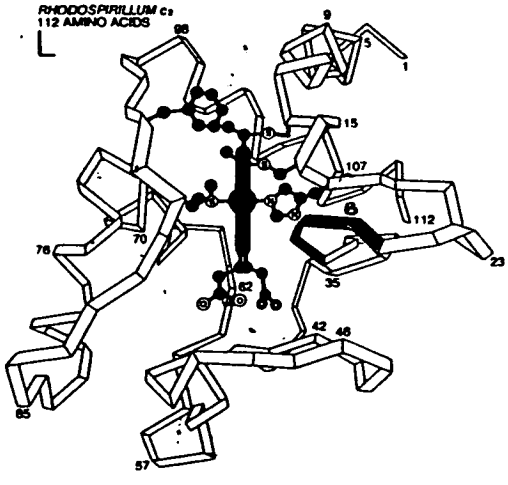
The proposed alignment of the algal sequence cytochrome c from *Anacystis nidulans* differs from the one proposed by Ludwig *et al.* (1982) on the basis of their crystallographic analysis. The difference is greatest in the third helix region and the following loop. In Ludwig's alignment this region is shifted to the right and the sequence, TTQVQN, is aligned with the helix breaker region, GKNA of *Pseudomonas aeruginosa* c551. While perhaps sensible when looking only at the sequence of the *An. nidulans*, it is more difficult to envisage when one considers homologous regions in some other algal cytochromes, such as in *Synechococcus*, MIQVQN and *Monochrysis lutherii*, VYQVTN. Furthermore, this region is the one that was least resolved in the electron density map. '*The left side loop (residues 50-60 in cytochrome c551 (Ps. aeruginosa)) is not completely defined in the algal c554 maps. The number of residues assigned to this loop depends on the identification of side groups in the preceding 40s helix. From the present protein densities , it appears possible to align either ile 44 or ile 47 with the leu 44 of c551. The former assignment would place nine residues in the 50s loop, and the latter only six, compared with eleven residues in (Ps. aeruginosa)c551 and twelve or thirteen in Chlorobium thiosulphatophilim c555*' (Ludwig *et al.*, 1982).



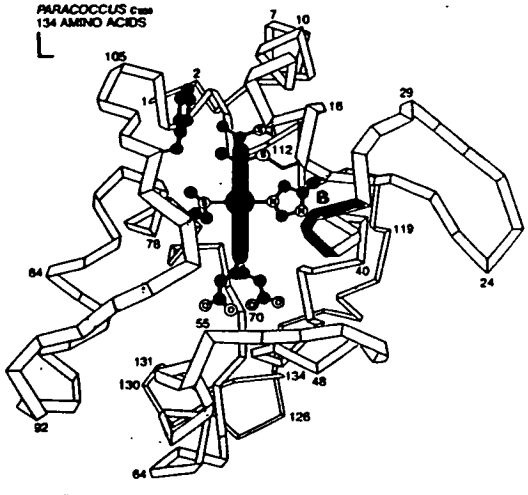
A. *Pseudomonas aeruginosa* c551



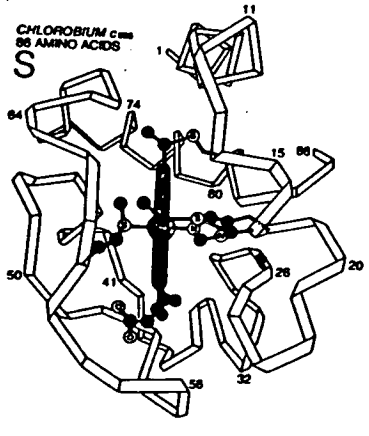
B.Tuna C



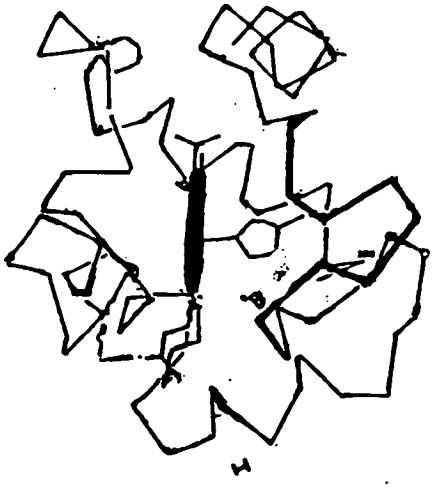
C. *Rhodospirillum rubrum* c2



D. *Paracoccus denitrificans* c550



E. *Chlorobium limicola* c555



F. *Azotobacter vinelandii* c5

Figure 6. 18. A schematic representation of six cytochrome c crystal structures.

A. *Pseudomonas aeruginosa* c551; B. Tuna c C. *Rhodospirillum rubrum* c2; D. *Paracoccus denitrificans* c550; E. *Chlorobium limicola* c555; F. *Azotobacter vinelandii* c5. A-E: After Dickerson (1980, F: Adaptaion after Carter (1985).

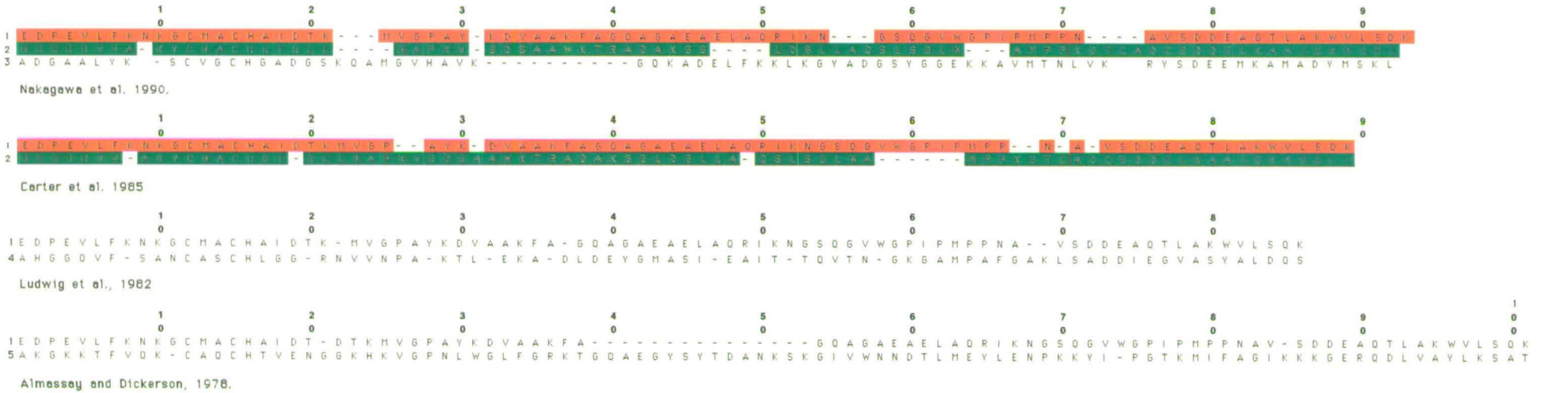


Figure 6.19. Alignments of cytochromes c based on three dimensional superposition of crystallographic coordinates.

1	*Pseudomonas aeruginosa*	c551
2	*Azotobacter vinelandii*	c5
3	*Desulfovibrio vulgaris*	c553
4	*Anacystis nidulans*	c553
5	Tuna cytochrome c	c (c2)

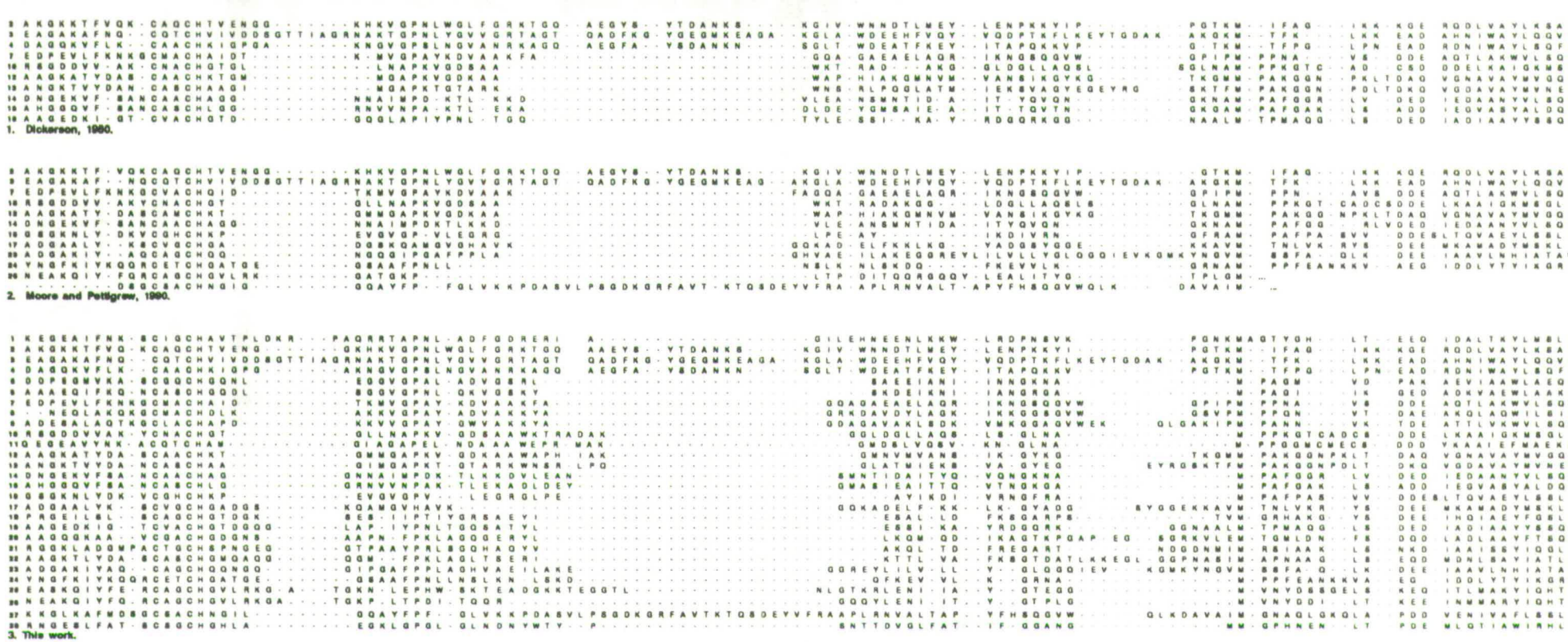


Figure 6.20. Four different overall alignments of cytochromes c.

The sequences are numbered according to the alignment of Hreggviðsson (this work).
The c2 sequnces are identically positioned in alignments 1, 2 and 3 . The Ps. aeruginosa c551 sequence is identically positioned in alignment 3 and 4.

1	*Bacillus PS3*	c(aa3)
2	Tuna cytochrome c	c2
3	*Rhodobacter sparoides*	c2
4	*Rhodopseudomonas acidophila*	c2
5	*Bacillus azotoformans*	c552
6	*Bacillus PS3*	c550
7	*Pseudomonas aeruginosa*	c551
8	*Hydrogenobacter thermophilus*	c552
9	*Rhodospirillum rubrum*	c553
10	*Azotobacter vinelandii*	c5
11	HiR	c5
12	*Chlorobium limicola*	c555
13	*Prosteocchloris aestuarii*	c555
14	*Porphyra tenera*	c553
15	*Anacystis nidulans*	c554
16	*Pseudomonas putida*	FC
17	*Desulfovibrio vulgaris*	c553
18	*Chlorobium limicola*	FC
19	*Paracoocus denitrificans*	c4
20	*Azotobacter vinelandii*	c4
21	*Azotobacter vinelandii*	c4
22	*Thibacillus neapolitans*	c4
23	*Thermus thermophilus*	c552
24	*Methylomonas capsulata*	c555
25	*Pseudomonas stutzeri*	cd1
26	*Pseudomonas aeruginosa*	cd1
27	*Methylobacter extorquens*	CL
28	*Pseudomonas aeruginosa*	CCP

1: Ishizuka et al., 1990; 2: Krell 1963, 1965 (cited in Moore and Pettigrew 1990); 3: Ambler et al., 1979; 4: Ambler et al., 1979; 5: This work; 6: Fujiwara et al., 1993; 7: Ambler, 1963; 8: Sanbogi et al., 1989; 9: Ambler, 1991; 10: Carter et al., 1985 (cited in Moore and Pettigrew 1990); 11: Ambler, 1991; 12: van Beeumen et al., 1976 (cited in Moore and Pettigrew, 1990); 13: van Beeumen et al., 1976 (cited in Moore and Pettigrew, 1990); 14: Ambler and Bartsch, 1976 (cited in Moore and Pettigrew, 1990);15: Margoliash (cited in Dickerson, 1980); 16: (MacIntire et al., 1985 (cited in Moore and Pettigrew, 1990); 17: Nakan et al., 1983 (cited in Moore and Pettigrew, 1990); 18: van Beeumen et al., 1990; 19: Ambler et al., 1978b; 20: Ambler et al., 1984; 21: Ambler et al., 1984; Ambler et al., 1986; 25: Smith et al., 1990; 26 :Smith et al., 1990;27: Hunn et al., 1988 (cited in Moore and Pettigrew, 1990); 28: Ronnberg et al., 1987 (cited in Moore and Pettigrew, 1990).

6.5.4. Grouping of Cytochrome C Sequences of Class I

A prerequisite for taxonomic ordering is that some hierarchical structure must be recognisable within the group and, perhaps also, that meaningful distances can be calculated between the units. Thus the question arises, is it possible to define equivalent taxonomic units of sequences in the cytochrome c family? Similarly, is it sensible, considering still limited sequence information, to define higher taxonomic units or subdivide further already defined subclasses. The IC and the ID subclasses may serve to illustrate this point. These subclasses appear not to be equivalent taxonomic units. The latter subclass may possibly be divided further, on the basis of identity, indel events and invariant and near invariant amino acids, into the *Rhodospirillum/ Rhodocyclus* sequences on one hand and the remaining ID sequences on the other (figure 6.5). In contrast, the algal cytochrome sequences and the c4 sequences, both of which now belong to subclass IC, may be different enough and internally consistent enough, each to merit a subclass status. It is clear that they have a common C-terminal sequence motif, but otherwise their similarity is slight. This similarity on the other hand may be used to group them together into a higher taxonomic unit. Group five, as defined below in section 6.5.4.2, is essentially the IC subclass excluding the algal sequences and some odd sequences, each of which may represent a subclass. Still, this group even by excluding the algal sequences, exhibits a greater sequence heterogeneity than the ID subclass alone. On the other hand, Group four (defined below in section 6.5.4.2) that consists of both the ID and the BacI subclasses, may perhaps be considered to be of equivalent taxonomic status to the c4 sequence group especially if some of the odd recently sequenced cytochromes are included as has been done below in section 6.5.4.2.

6.5.4.1. Grouping of Class I cytochromes c into subclasses

Pettigrew and Moore (1987) point out that sometimes differences in size reflect only minor genetic changes in otherwise related proteins. Thus, the c2 cytochromes form a homogenous group with more than 30 % identity, whereas on the basis of size, the group can be split up into two (Ambler 1982, Dickerson, 1980). These differences involve short loops on the surface of the protein. When defining a sequence subclass or any other taxonomic unit, it has to be clear what is a minor genetic variation. The variation must be to such a degree, that it can be said to have shifted

the structure "noticeably" from one point to another in sequence space. In general high % identity suffices to delimit their sequence subclass. Some minimum pairwise identity values of subclasses are given in table 6.2.

The following proposal is made as to which cytochrome sequence groups may be considered as having an equivalent taxonomic status of a subclass. It basically follows the scheme of Ambler with some suggested revisions:

1) The c2 sequences, which are divided into medium and large cytochromes by Dickerson (1980) and IA and IB subclasses by Ambler (1980).

2) The cytochromes c(aa3) from *Bacillus* PS3 and *Thermus thermophilus,*

3) The ID subclass, which are *Pseudomonas* c551 like sequences,

4) The BacI sequences from *Bacillus* species,

5) The c5 sequences

6) The c555 sequences from green sulphur bacteria

7) The algal sequences and

8) The c4 sequences.

The two cd1 sequences (Smith and Tiedje, 1990), and the remaining odd sequences may be considered as the only sequenced members of still other subclasses. An attempt is made below to define a higher taxonomic unit than subclass, the group.

6.5.4.2. The grouping of cytochrome c subclasses

By reference to the procedure of taxonomic ordering (Sneath, 1989) of protein sequences as it is defined in section 6.5.1, the alignment of the cytochrome sequences can be considered as an attempt to establish sameness. The alignment (6.17a) shows that the sequences converge on a pattern of amino acids of similar properties, which can be defined as the signature of cytochromes c of Class I (figure 6.17a.b. and section 6.5.3.1.3). The sequences in the alignment are, clearly, highly divergent and this is reflected in the identity matrix (figure 6.17c). Although some clusters can be discerned, most of the identity values are far too low to allow the taxonomic structure of this protein family to be carved out. It follows that different criteria must clearly be used to establish sameness quantitatively at different taxonomic levels.

The identity values between sequences of different subclasses and the odd unclassified sequences are very low (figure 6.17b. These low values are at the level where specific affinities between sequences are obscured

by background noise. Paradoxically, background noise includes invariant amino acids such as the cysteines and the histidine of the haem site, and the sixth ligand methionine. These residues make a significant contribution to the identity values, because of the small size of the sequences and the considerable divergence between them. Invariant or redundant traits will skew any calculations of similarities or distances for taxonomic purposes as they do not provide any discriminatory information. Such traits are, for example, excluded in numerical taxonomy analyses of bacteria.

In the cytochrome c sequences there are also sites or small regions that are semi-invariant between all subclasses. These sites or positions are occupied by one of a set of relatively few amino acids in all cytochrome c sequences of Class I. The amino acids appear to be interchangeable at these positions within subclasses and also between them. A striking example of such region is the acidic region in the C-terminus. It consists of three adjacent positions where there is a predominant occurrence of either glutamate or aspartate and to a lesser extent glutamine and aspargine. A less conspicuous region of this kind, is between the two haem site cysteines, where there is a predominance of small aliphatic amino acids (alanine, serine, glycine). Such regions may occasionally increase pairwise identity values, so that unwarranted assumption about specific affinities can be made. Secondary structural elements, whether regular core elements, turns or coil regions, may show similar amino acid patterns. These structures may be homologous, but they can also be examples of convergent evolution. Lastly, at very low levels of identity, as is observed between many of the cytochrome c sequences, coincidental positional identity can have obscuring effects on the score.

The following grouping of the sequences is based on the alignment in figure 6.17a and the corresponding identity matrix cast into a dendrogram by the UPGMA method in figure 6.17b.

Any scoring method which reduces sequence similarities to single number scores, only holds a part of the information actually contained in the sequences. The grouping is therefore scrutinised or evaluated by visual inspection of the alignment. To enhance the similarities and difference of the sequences in the alignment (figure 6.17a), the individual amino acids have been coloured according to their physicochemical character. The most important factors thus revealed are the apparent group specific sequence motifs, hydrophobic amino acid patterns, the

number and positions of gaps and resulting demarcation of loop and core regions.

Group 1: This group consists of the cytochromes c2, which have been divided into two subclasses IA and IB, by Ambler (1977, 1982, 1991), and the apparently homologous *Bacillus* PS3 cytochrome c(aa3) and the *Thermus thermophilus* cytochrome caa3. The conserved sequence characteristics include; a near invariant phenylalanine in position seven, a conserved hydrophobic residue in position 18, a near invariant glycine in position 39, an aromatic amino acid in positions 55 (if histidine is included) and 63, an invariant proline in position 67 and a near invariant lysine in position 85. The c(aa3) sequences have a similar positional hydrophobicity pattern as the c2 sequences. The indel events proposed, to maximise identity are few and only in the loop regions. The dipeptide insertion from position 87 to 88 is found also in c2 sequences from *Rhodspirillum fulvum* (figure 6.3) and *Rhodospirillum molischianum.* Perhaps the most noteworthy is conservation of the region preceding the sixth ligand methionine between the c2 sequences and the *Bacillus* sequence including the proline residues, in positions 67 and 82. This sequence part is very characteristic of the c2 (and c) sequences. The *Bacillus* sequence is least similar to the other sequences in the C-terminal region after the sixth ligand methionine. Subclass IB can be divided further, on the basis of size differences in the loop regions. The similarity of the c(aa3) cytochrome sequences to the c2 cytochromes is striking and has interesting evolutionary implication (sections 6.6 and 6.7). This sequence type can be defined as Small cytochrome c2 in the terminology of Dickerson (1980). Interestingly, it appears that the differences of the caa3 sequences to the c2 cytochromes involve insertions (and a deletion of one amino acid right after the sixth ligand methionine) in the c2 sequences. The latter basically conforms to the structural blueprint of a small bacterial cytochrome c.

Group 2: This group consists of the ID subclass and the BacI sequences from the genus *Bacillus.* Arguments for their relatedness or similarity have been given in section 6.3.4. The *Bacillus* sequences have relatively high scores with the algal sequences. Thus the PS3 sequence has 34% identity with the *Anacystis nidulans* cytochrome c553. This is not indicative of any special relatedness or even structural similarity in my opinion, for the following reasons: 1) The hydrophobicity pattern is considerably different, both positionally and in identity. It is more similar

between the BacI sequences and the ID subclass in both aspects. 2) A major contribution to the identity score comes from regions or positions, which are not important in defining the subclasses themselves. Either these are highly variable positions within the subclasses or they are conserved, but the conservation is not restricted or characteristic of either subclass. These residues include the conserved haem cysteines and the haem ligand residues, residues in the acidic region and between the haem cysteines, and up to five residues in the loop region between the third helix and the sixth ligand site, which probably takes up a β-turn conformation in both proteins. In such turn regions a high similarity may be expected. 3) The distribution of charged amino acids is different (figure 6.5).

Group 3: This group consists of two sequence groups, the c5 cytochromes, subclass IE (Ambler 1991), and the c555 cytochromes from the green sulphur bacteria *Chlorobium limicola* and *Prosthecochloris aestuarii*. Despite the very distant relationships of the bacteria containing the Group three sequences, the c5 and the c555 sequences respectively, the identity is relatively high. Also, the hydrophobicity pattern in core element regions is generally conserved. The main difference between the sequences is an insertion of a number of amino acids between the third helix region and the sixth ligand methionine of the green sulphur bacteria cytochromes. Apart from this only one minor indel event of one amino acid needs to be proposed to maximise identity in loop two. A conserved pair of cysteines is also found in the C-terminal region of the c5 sequences, closely following the sixth ligand site. Until recently, no sequenced c5 sequences had an aromatic acid in position 107 at the C-terminus, which is characteristic of all other cytochromes of Class I. In the latest sequence published (Ambler, 1991), from a strain of uncertain phylogenetic status H1R, a bulky aromatic amino acid, phenylalanine, is found in this position.

Group 4: In this group are the algal sequences, presented in the alignment by the *Porphyra tenera* and *Anacystis nidulans* c555 cytochromes, which are grouped in the IC subclass by Ambler (1991). They have an identity of 58 %. The algal cytochromes show a considerably high identity (a minimum of 50%), even though some are from cyanobacteria and others from chloroplasts of eukaryotic algae. Perhaps one other sequence in the alignment can be included in this group, the flavo-cytochrome c from *Pseudomonas putida*. It has a similar hydrophobicity

pattern to the algal sequences, with considerable identity in parts. The similarity is greatest in the C-terminal region and identity is high. Noteworthy is the identity in the sixth ligand region, including a conserved proline in position 87 and a phenylalanine in position 89. Two indel events have to be proposed, which slightly mar the relationship, one in a core region immediately after the conserved proline in position 32 and another in the second loop region. Otherwise, an almost identical delimitation of secondary structure elements is observed. The glycines in the second helix region also indicate that this sequence part may have a different conformation from the corresponding region in the algal cytochromes.

If this is a historically correct alignment, some major differences would be expected, considering the great phylogenetic distance between the cyanobacteria and the genus *Pseudomonas*. However, as has been stated before, a taxonomic grouping of sequences, based on sequence characteristics, is not necessarily the correct historical alignment and the similarity is not particularly strong.

Group 5: This group consists of the c4 cytochromes from purple bacteria, besides many of the previously unclassified sequences, the flavocytochrome c sequence from *Chlorobium limicola*, the *Thermus thermophilus* c552 cytochrome and perhaps more doubtfully the *Pseudomonas* cd1 sequences and the *Methylobacter capsulata* cytochrome c555. This grouping is based on the specification for the IC subclass, which was laid down by Ambler (1982). The signature of the c4 cytochromes, which are central to the subclass, serves as the template for the group, to which the other sequences adhere to a varying degree. Sequence characteristics are the C-terminal sequence motif, -M-----LS---I----Y-, and aromatic and proline residues after the haem binding site, preceding the proline, which hydrogen bonds to the fifth ligand histidine. Most of the sequences also have glycine next to the fifth ligand histidine and an aromatic amino acid is commonly found in position 64 of the third helix. An indel event has to be proposed in the third helix region to maximise identity to sequences of other groups. The *Chlorobium limicola* flavocytochrome c sequence does not have the second proline between the haem site and the proline site, but the demarcation of the core elements is almost identical and the hydrophobic pattern is very similar. The affinity in terms of identity has been noted by van Beeumen *et al.* (1990). The *Methylococcus capsulata* sequence and the cd1 sequences have the

greatest affinity to the *Thermus thermophilus* c552, which can perhaps be considered as the intermediate sequence between the apparently two forms of sequences put into the group.

The *Thermus thermophilus* c552 cytochrome may perhaps have the most ancestral characteristics of all cytochromes c sequences of Class I. The Group five is also the most heterogeneous group and its members apparently have widely different functional roles. Perhaps it can be claimed that the prototype Group five cytochrome in its basic ancestral form, which still can be discerned in its descendents to a varying degree, may have represented a relatively unspecialised electron transport protein. At different times it may have been recruited for more specialised adaptations with resultant structural changes and shift in sequence space. The more ancient offshoots may be represented by members of the other cytochrome c groups.

Odd sequences: The remaining sequences of the alignment cannot be confidently placed. The *Desulfovibrio vulgaris* cytochrome c553 is strikingly different from other cytochrome sequences. The most notable difference is the absence of proline in position 32, which in most other sequences hydrogen bonds to the fifth ligand histidine. The sequence has relatively high identity to the *Chlorobium limicola* sequence and perhaps slight affinities to other fifth group, but the hydrophobicity pattern is, dissimilar both positionally and in identity. The *Methylobacter extorquens* sequence and the cytochrome c peroxidase sequences, while clearly belonging to Class I cytochrome c sequences, have the slightest similarity to each other, but even less to any other sequences. Their taxonomic status is obscure.

Cytochrome c sequences from the genus *Bacillus*: Considering the great phylogenetic distance, the *Bacillus* sequences are surprisingly similar to cytochrome sequences found in bacteria from the phylum of purple bacteria. The marked similarity of the *Bacillus* PS3 cytochrome c(aa3) and *Thermus thermophilus* cytochrome c(aa3) to each other and their great affinity to the c2 sequences, is especially unexpected, and perhaps indicates that the electron transport pathways in these bacteria are more related than previously expected. The P3&P4-c555 sequence from *Bacillus azotoformans* was only partially sequenced and cannot be classified with confidence. The sequence has strong affinity with the c4 sequences of the IC subclass, in the sequence region covering the haem

site, the first loop and the proline that hydrogen bonds to the fifth ligand histidine, but also to the BacI sequences immediately after the fifth ligand histidine. The similarity to the c4 sequences at the C-terminus is weak, but no greater similarity is found to sequences in any other subclasses. The placement of the large peptide in the middle of the protein in figure 6.3 is highly conjectural. Perhaps, it can be said that this cytochrome c sequence forms a subclass of its own, but that it is has strong affinity to the Group five sequences. A further analysis, whether supporting this or not, must await the full sequence.

6.5.4.3. Discussion of the proposed groups

It is clear that a great many of the similarities between the cytochrome c sequences in the identity matrix (figure 6.17c) are statistically insignificant, and while the branching order in the dendrogram is perhaps not utterly incorrect, especially at lower taxonomic levels, the branching order of the major groups must remain highly doubtful. These groups and the remaining unclassified sequences are more or less discreet units. There are no sequences which can be considered transitional between them two groups, except for the c(aa3) cytochrome sequences from *Bacillus* and *Thermus,* which have the amino acid signature of a cytochrome c2 but are structurally much more like small bacterial cytochromes than the previously sequenced c2 sequences. It is also to be expected that with more sequences available, more subdivision will be possible of the already convincingly defined cytochrome c groups.

As said above, different criteria must clearly be used to calculate distances or similarities for classification purposes, at the different taxonomic levels. At the lower levels amino acid identity is sufficient, but at higher levels it is perhaps not. Further resolution may be obtained with weighted amino acid scoring matrices instead of using positional identity (Dayhoff *et al.*, 1983). In the above analysis, beside the identity values, specific sequence motifs, invariant amino acids, indel events, hydrophobicity patterns were used. These sequence traits are quantifiable within a specific group by the profile method of Gribskov, and can be represented graphically for visual inspection as was done in figure 6.8 for the ID and the algal subclasses, but relationships between different sequence groups based on these traits are much more difficult to quantify. An interesting approach, which I am not aware has been looked into, is to calculate pairwise similarities or affinities between sequence groups themselves rather than between their individual constituent sequences. A

subclass or any sequence group for that matter, can be reduced into a numerical group-specific profile matrix. The profile consensus sequences can then be used as parameters of similarity between sequence classes, and pairwise cumulative scores calculated from the profile matrices (see appendix B). In a protein profile matrix some sequence positions are enhanced in importance and others devalued. Thus, similarity should have a greater chance of being significantly detected and meaningful distances calculated. This requires that a great number of sequences has to be available, which limits the usefulness of the method. Figure 6.21 shows a rudimentary affinity matrix obtained by pairwise scoring of the consensus sequences of the different subclasses in each others profile matrices (see appendix B).

		1	2	3	4	5	6
1	Algal	1,00					
2	Bacl	0,46	1,00				
3	C4	0,46	0,43	1,00			
4	C5	0,26	0,38	0,48	1,00		
5	ID	0,42	0,64	0,45	0,43	1,00	
6	C2	0,41	0,48	0,45	0,30	0,39	1,00

Figure. 6.21. Affinity matrix between 6 subclasses of cytochromes of class I.

The affinity scores converge at around 0.40 (lowest score=0, highest score=1) and perhaps this indicates that hierarchical taxonomic structure above the level of Group (as it has been defined in this thesis) is an insurmountable problem, perhaps because of the very small size of the cytochromes. The only marked affinity above this level of approx. 0.4, is between the ID and the Bacl subclasses (0.6).

Even though the lineage of cytochromes c has not been unravelled, perhaps the signatures of the Class I cytochrome c sequences indicates,which are the most ancient sequence characteristics. In other words, one can perhaps discern the idea of cytochrome c, of which the extant sequences are but 'incomplete reflections'.

6.6. DISTRIBUTION OF CLASS I CYTOCHROMES C IN EXTANT EUBACTERIAL SPECIES

In the following discussion, the taxonomic designation, of Ambler (1977) (subclasses IA-IB), Dickerson (1980) ((Small(S), Small*(S*), Medium(M) and Large(L)) and the traditional ones, c1-c8, (the numbers are usually subscripted, but not in this thesis), adopted and proposed by Pettigrew and Moore (1987), will be used interchangably. These designations are defined in Tables 6.2, and 6.3. the ones proposed in this thesis, c2'-c7', (proposed homologues to the corresponding c2-c7 cytochromes) are found in table 6.3.

Cytochromes of Class I are important or essential constituents of many of the major metabolic pathways. They are found in phylogenetically very diverse eubacterial species. Their great variety, which is manifest both within and between species, and its wide distribution indicates that it is a very ancient protein. Evolution must have had time to spawn all these diverse structures.

Within a species, gene duplications may have enabled sequence divergence and corresponding adaptation for different functions. A recent event of such kind may be the duplication of the BacI sequences in Gram positive *Bacillus* species. Such duplication may ultimately result in specialised three dimensional structural modifications, or a shift in sequence space and the emergence of a different subclass. Another explanation for this diversity is that lateral transfer has brought together cytochromes which originally evolved in separate species.

Until the publication of the cytochrome c sequences from the genus *Bacillus* in recent years, bacterial cytochromes c of Class I had only been sequenced from bacteria belonging to four of the ten major subgroups of eubacteria or phyla, as they are defined by Woese (1987) The great majority of the Class I sequences are confined to a restricted range of bacteria. Most are from the alpha, beta and gamma subgroups within the phylum of purple bacteria. Two sequences here have also been reported from*Desulfovibrio* in the delta subgroup of the purple bacteria. Two cytochromes c, from *Thermus thermophilus,* have been sequenced from the *Thermus*/Deinococci group, two have been sequenced from the green sulphur bacteria and approximately a dozen have been sequenced from the phylum of cyanobacteria. Figure 6.22 shows how the various groups of cytochromes c types (designations c1-c8 and c'-c6', defined in tables 6.2 and 6.3) map onto the phylogenetic 16sRNA tree. This figure is interesting

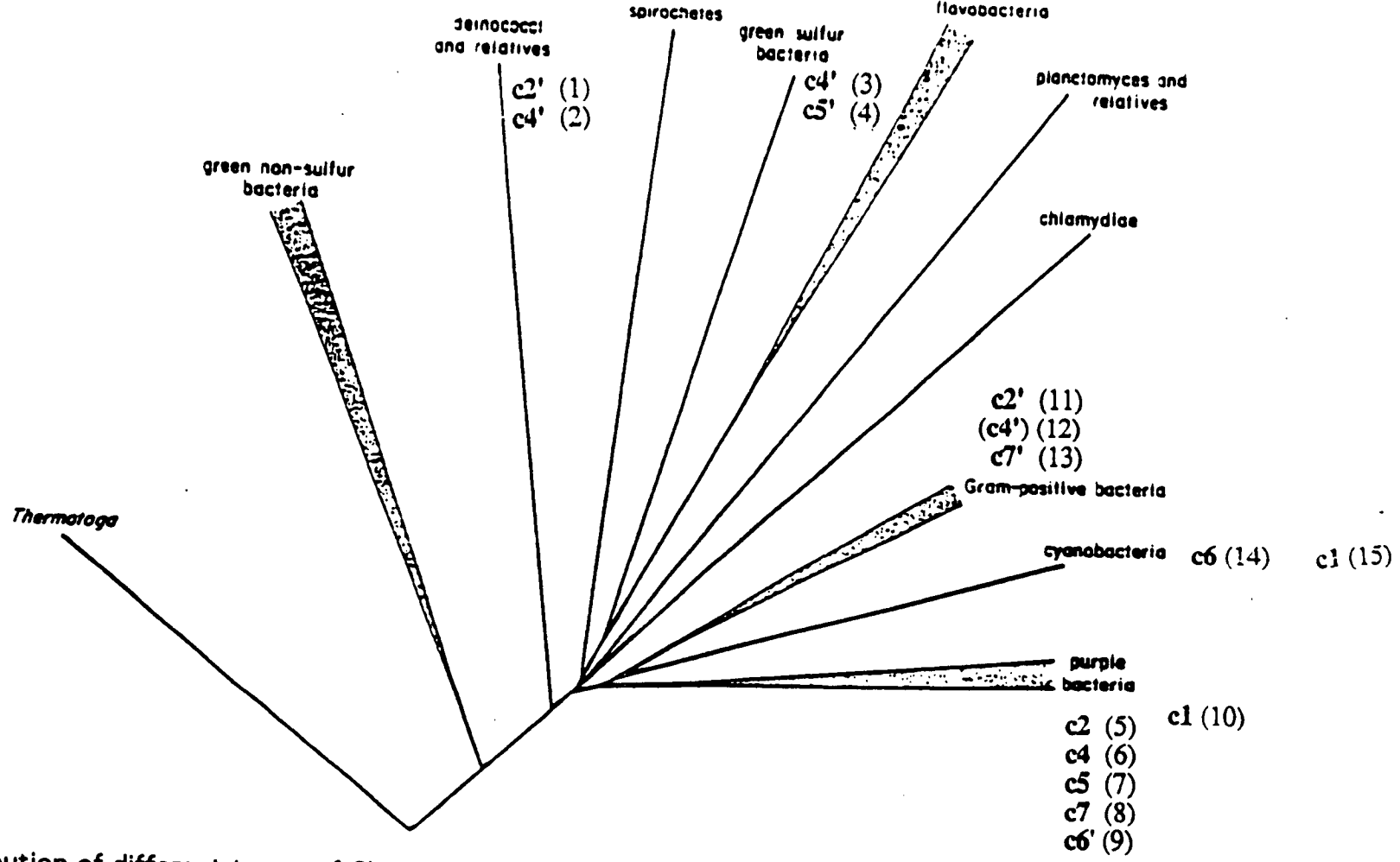


Figure 6.22. Distribution of different types of Class I cytochromes c and c1/f6 cytochromes in the 16sRNA eubacterial phylogenetic tree.

The 16sRNA tree is reproduced from Woese (1987). c2 , c4, c5, c7 (subclass ID) and c6 (algal, soluble f) are designations in the approved nomenclature system or cytochromes c or proposed by Moore and Pettigrew (1990). c2' c4' c5' c7' and c6' are designations used in this thesis and refer to proposed structurally analogous or homologous cytochromes c previously unclassified or reclassified in this work. Same numbers of cx and cx' means that they belong to the same homologue group. c2 The cytochrome c. The groups are defined in section 6.5.4.2. in table 6.3 and the alignment in figure 6.17a. c2 and c2' belong to group 1 c4 and c4' to group 5, c5 and c5' to group 3, c6 and c6' to group 4 and c7 and c7' to group 2.

(1) *Thermus thermophilus* c(aa3); (2): *Thermus thermophilus* c552; (3): *Chlorobium limicola* cytochrome CF (4): *Chlorobium limicola* c555; (5) cytochrome c2 purple bacteria; (6) cytochrome c4 purple bacteria, (7) cytochrome c5 purple bacteria(8) the ID subclass purple bacteria (9) *Pseudomonas putida* FC or Ectothiorhodspira halophila c551 (Meyer 1991, see also Taylor 1989); (10) cytochrome c1 (an f6 homologue) found inpurple bacteria and mitochondria, (11) *Bacillus* PS3 caa3 cytochrome c, (12) *Bacillus azotoformans* P1-c555 (a tentative affinity, only a partial sequence) (13)*Bacillus azotoformans* P1-c552, (14) algal (soluble f) cytochromes (15) cytochrome f6 (and c1 homologue) found in cyanogen bacteria and chloroplasts.

for two related reasons:

-Firstly, *Thermus thermophilus* contains cytochromes c. *Th. thermophilus* belongs to the apparently 'ancient' group of*Thermus*/Deinococci. It branches earlier from the tree than all the other phyla, from which cytochromes c have been sequenced. It is therefore interesting to see if it contains recognisable forms of cytochromes of any of the groups already defined for cytochromes c from bacteria in other phyla, or how far back into the ancestry of eubacteria one can trace sequence traits of extant subclasses (presuming that lateral transfer has not 'corrupted' things). Two cytochrome c sequences have been reported from this ancient group, a cytochrome c552 from *Thermus thermophilus* (Titani *et al.*, 1985) and a cytochrome c which is fused to a cytochrome c oxidase (caa3) (Mather *et al.*, 1991). In section 6.5.4 it is argued that the former may belong to the same homology group as the c4 cytochrome c sequences of purple bacteria and that the latter is homologous to the *Bacillus* PS3 cytochrome caa3 and in a homologue group with the c2 sequences from mitchondria and purple bacteria.

-Secondly, sequences from the same homologue group, even if certainly highly divergent, are found in more than one phylum. c2 cytochromes have previously only been found in purple bacteria and mitochondria, but as argued above, recognisably homologous sequences appear to be present in the phyla of *Thermus*/Deinococci and Gram positive bacteria. Similarly, sequences structurally analogous or homologous to the c4 cytochromes in purple bacteria, can be found in the phyla of *Thermus*/Deinococci and also in the green sulphur bacteria. They are represented, respectively, by the cytochrome c552 from *Thermus thermophilus* and the flavocytochrome c from *Chlorobium limicola*.. Gram positive bacteria may contain a c4 sequence homologue, the cytochrome P3&4-c555, which is found in *Bacillus azotoformans* (this work), but the evidence for this is still tentative. In this thesis it has furthermore been argued that Gram positive bacteria contain cytochromes c (the BacI cytochromes) homologous to the ID sequences found in species of the beta and gamma subdivisions of the phylum of purple bacteria . In addition the cytochromes c5 in purple bacteria appear to have homologues in bacteria in the phylogenetically distant phylum of the green sulphur bacteria. These homologues are represented by the c555 of *Chlorobium limicola* and the cytochrome c555 from *Prosthecochloris aestuarii.*

At least three explanations can be put forward for the high diversity but still, apparently similar cytochromes in very phylogenetic diverse

bacteria. Any one explanation is perhaps not satisfactory and all three may fill in a part of the picture. They are:

1) **The same or similar cytochrome structures or sequence types, which are found in such phylogenetically distant species, are examples of convergent evolution, and the observed variety within species is a consequence of internal duplications.** While convergent evolution of sequence parts is perhaps common, a true convergence of overall structure is perhaps unlikely to occur. In sections 6.5.4.2, arguments have been given for the structural similarity of the different group members. The similarity does not necessarily have to be realised in three dimensions, for example, between the BacI cytochromes and the ID cytochromes from purple bacteria. So, despite primary sequence similarities the more divergent members of any group may specify entirely different messages or structures and be the result of slow (superficially?) convergent evolution in the different lines of descent.

2) **The main subclasses were established before the divergence of the majority of the eubacterial phyla and representatives may be found in any of them.** It is perhaps unlikely that such a structural variety as is observed among extant cytochromes c, was established at such a very early stage of bacterial evolution. The 16S rRNA tree would imply that this would have occurred before the divergence of most of the phyla. However, genetic mechanisms may have been less restrictive or less conservative at these times, so that incidences of duplications may have been much more common. Also, proofreading apparatus might have been much less effective so that many more structures were tried out and tried much faster on a geological timescale. The extant cytochrome structures may then be only the very few successful of many more disastrous ones.

3) **Lateral transfer has played an extensive role in the distribution of the various cytochrome c structures.** It is not known how common or how important lateral gene transfer has been for the evolution of metabolic pathways (see chapter 1) Even if it can be argued that cytochromes from different phyla belong to the same homology group, which could be explained by lateral gene transfer between phyla, the divergence is to such an extent, that such an event

must have taken place in the far ancestry of extant species. On the other hand, lateral transfer, while may be infrequent, may have wrought dramatic changes to the future of eubacterial lines, and perhaps different from the one originally written into their genes. This may never be proven, but perhaps the existence of the aerobic lines of descent in the phylum of Gram positive bacteria, reflects such an incidence. The phylum is, according to the old interpretation of its metabolic constitution, anaerobic in origin (Dayhoff *et al.* 1978). This is also the conclusion of the more recent studies of Woese (1987) based on analyses of 16S rRNA sequences, where he proposes that aerobiosis has risen independently in the phylum (see below)

6.7. THE EVOLUTION OF METABOLIC PATHWAYS IN THE GENUS BACILLUS AND THE STATUS OF GRAM POSITIVE BACTERIA IN THE EUBACTERIAL KINGDOM

6.7.1. Phylogeny and Evolution of Metabolic Pathways

The importance of 16s RNA analyses for establishing relationships and providing at least to some approximation, an arbitrary time course of bacterial evolution is evident. However, the information content of rRNA is limited and to be able to map out the crucial events of evolution onto the phylogenetic tree, thorough studies of other macromolecules and self contained bacterial systems must be undertaken. In this sense the comparative study of the electron transport chain must be important and a necessary counterpoint to the phylogenetic studies.

What has emerged from comparative studies of the different metabolic groups of bacteria, based on functional and physicochemical studies, is that there appears to be more or less, an ancient core of distinct redox carriers common to eubacteria across the metabolic spectrum and also intermediate ones common to major radiations or more confined lines of decent (Jones, 1985, Anraku, 1988).

Ambler (1991) argues that the most plausible explanation for the sporadic occurrence of the same complex metabolic functions such as denitrification, methylotrophy, nitrogen fixation besides aerobiosis and photosynthesis, in most of the major eubacterial phyla is due to the occurrence of lateral transfer else the primitive organism would have had the capability of carrying out all these functions and the evolution would have been a process of regression. In contrast, Stouthamer (1992) postulates that a number of the complex metabolic systems (nitrification, denitrification, aerobiosis) arose earlier than thought and any one may have been lost numbers of times in different lines of descent and furthermore, that lateral transfer has played a minor role. Traditionally, photosynthesis is placed central in the discussion of evolution of the metabolic pathways and it is argued that many of the respiratory pathways are but offshoots of that system (Dayhoff *et al.*, 1978; Jones 1985; Woese, 1987). This may be the case indirectly as photosynthetic species are found in the second deepest eubacterial branching and as the necessary supplier of oxygen, but perhaps photosynthesis is also the wild card. The evolutionary scenario might become more sensible if the distribution of photosynthesis could, in some instances, be explained by lateral transfer.

Meyer (1991) has argued along these lines, on the basis of the distribution patterns of cytochromes c and reaction centres. He, furthermore, argues that the metabolic substructure of the *Pseudomonas* line of descent,is ancestral to photosynthesis which is supposed to predate the major radiation of eubacteria (Woese, 1987) (see next section).

Common physical properties do not necessarily reflect any evolutionary relationship, as exemplified by the different cytochrome classes. Also, some components of the electron transport chain in different lines of descent, albeit of the same protein family, might have acquired the same role or function in different bacteria by entirely different routes. To be able to trace the history of metabolism there have to be means to distinguish between superficial and true resemblances and this can be achieved by comparative sequence studies of particular components of the electron transport chain. The choice of protein for evolutionary studies must ideally be important for the particular organismal trait under study. For example terminal oxidases might reveal something about the origin of aerobiosis, whether it were mono or multiphylogenetic. Similarly, nitrate reductase sequences might tell something about the evolution of nitrate respiration. The more information on all types of proteins from the same pathway, the more reliable the conclusions can be drawn. Perhaps, instead of only cladistic interpretation of sequence information, evolutionary signatures based on patterns of proteins of known structures, such as the different cytochrome c variants, can be defined for specific metabolic adaptations and used for evaluation of relationships. This will help to make the 16S rRNA eubacterial tree, perhaps the only frame of reference to be had, historic as well as just phylogenetic. It may also reveal, to what extent the different cellular apparatus and their constituent proteins follow the same organismal lineage.

In the above context it is interesting to compare the distribution patterns of the various cytochrome subclasses in purple bacteria subgroups with the pattern in the genus *Bacillus*. The alpha subgroup of the purple bacteria has c2 cytochromes, but none of the small ´pseudomonas´ cytochromes, the c5, c4 or ID subclass cytochromes, which are found in the beta and gamma purple bacterial subgroups. Species of latter subgroups appear, on the other hand, to do without cytochrome c2 (Bartsch, 1991). If the structural similarities of the *Bacillus* cytochromes to certain purple bacterial cytochromes reflect true homology, then c2 and ID homologues, (the caa3 cytochrome and BacI cytochromes) coexist in

Bacillus species in contrast to what has been observed in the purple bacteria. Even the P3&4-c555 cytochrome may also be a homologue to a purple bacterial cytochrome structure, the c4 sequences, but the extent of their affinity is still obscure. The implications of this distribution of cytochromes c for the evolution of metabolic pathways in Gram positive bacteria is discussed in the next section, 6.7.1.

Cytochrome c can be considered a rather unspecialised protein able to take on many guises. It has certain characteristics common to co-enzymes in its basic role in the electron transport chain (Salemme, 1977). In any one form, it is perhaps optimally adapted to a special function, involving a pair or a very limited number of redox 'partners', but in special circumstances it can substitute for other electron transport proteins (Anraku, 1988; Pettigrew and Moore, 1987; Moore and Pettigrew, 1990). On the other hand, cytochromes c are at the low-free-energy end of the metabolic pathways and may have been fundamental or important for evolution of specialised metabolism in eubacteria. This may be indicated by the many structural varieties of cytochromes c, which are often found to co-exist in species of great metabolic versatility. The functional flexibility on one hand and the inter and intraspecies diversity of cytochromes c on the other, is a controversy that remains to be explained.

6.7.2. Evolutionary History of Electron Transport Pathways in The Gram Positive Phylum

The phylum Gram positive bacteria poses some interesting evolutionary questions which highlight the difficulties of evaluation and interpretation of phenotypic as well as sequence data whether it be from rRNA or proteins. It emphasises the need to put different data into context.

Considerable metabolic versatility has been found among Gram positive bacteria (Claus *et al.*, 1986). Apart from their "major" pathways fermentation, and aerobiosis, photosynthesis (*Heliobacterium chlorum*), nitrogen fixation (*Bacillus* and Actinomycetes), sulphate respiration (*Desulfotomacculatum*), denitrification (*Bacillus* and Actinomycetes), methylotrophy (*Bacillus*), all have been found in extant Gram positive species. These same metabolic phenotypes are found in a variety of species outside the Gram positive phylum and this poses the question of monophyletic versus polyphyletic origin of these pathways and the possibility of lateral gene transfer (see preceding section) (Ambler, 1991;

Stouthamer, 1992). When the component composition of electron transport pathways from a range of Gram positive bacteria is analysed on the basis of biophysical properties, it seems that one can find evidence to support any evolutionary scenario proposed (see section 6.1.2 and Jones, 1985).

Dayhoff *et al.*, (1978) attempted the phylogenetic reconstruction of eubacteria by making a composite tree based on cytochrome c, 5sRNA and ferredoxin sequence analyses. In this tree, the clostridia have a position at the base of the tree. This accorded with general ideas of the evolution of metabolism, which proposed that fermentation was the earliest metabolic trait. Therefore clostridia were considered to have retained this ancient if modified metabolic pathway.

The early catalogue analysis of 16S rRNA Stackebrandt and Woese (1981) also suggested that the ancestral phenotype was anaerobic and fermentative. Thus the deepest branching genera of the clostridial and actinmycetal subdivisions are fermentative, and clostridia are found in most sublines of the presumably more ancient clostridial subdivision. It is assumed that the more ancient the trait, the wider distribution it will have in extant distantly related species. This also implies that aerobiosis has evolved independently in the Gram positive phylum and furthermore independently in the clostridial and actinomycetal subdivisions (Stackebrandt and Woese, 1981; Woese, 1987). Although not explicitly stated, this conclusion does not seem to have been contradicted by 16S rRNA signature analysis (section 1.1.5.1), which should have revealed if the deep branching were due to faster evolutionary rates in the fermentative species. Woese (1987) has subsequently proposed that the *Bacillus* subline within the clostridial subdivision developed aerobiosis concurrently with the development of aerobic conditions on earth (Woese 1987). Thus *Lactobacillus* is anaerobic, *Streptococcus* microaerophylic, and *Bacillus* aerobic, and this accords with the branching order of these genera. However, the existence of cytochromes c and cytochrome oxidases aa3 in the *Bacillus* subline seems to necessitate a preadapted electron transport chain more fully developed than the rudimentary one of *Lactobacillus* and the extant *Clostridium*, or else it would have no antecedents in the phylum and closer relatives in other phyla.

The comparatively recent discovery of the Gram positive, strictly anaerobic, photosynthetic species, *Heliobacterium chlorum* may hold the key to the evolutionary history of the phylum. The available data (Woese *et al.*, 1985,) indicates that it is the sole member of a moderately deep branch in the clostridial subdivision close to the bacillus group (85% 16S

rRNA sequence similarity with *Bacillus subtilis* (Woese *et al.*, 1985). The components of its electron transport system have not yet been extensively studied, but the reaction centre seems to be unrelated to ones previously studied and studies indicate at least three different cytochromes c (Clinton-Fuller *et al.*, 1985). On the basis of its discovery Woese *et al.* (1985) have proposed a photosynthetic origin for the Gram positive bacteria as it is unlikely that such a complicated trait has evolved independently in the Gram positive phylum. Furthermore, they postulate that it is also anaerobic because of the metabolic substructure of the phylum. The former, if correct, and therefore also the finding of *Heliobacterium chlorum*, can be said to have been anticipated by Jones (1985), by analysis of the physicochemical characteristics of the respiration pathway components in the genus of *Bacillus*, on the basis of which he suggested that a photosynthetic electron transport chain might have been the ancestral system

By including outside species to a particular taxonomic group, the phylogenetic tree obtained by sequence analyses, can be rooted and the branching order determined. Consequently, such a tree should expose fast evolving lineages. Unfortunately few 16S rRNA sequences have been published so far from the Gram positive phylum, but recently a rooted dendrogram (figure 6.2.3), based on complete 16S rRNA sequences from a greater variety, if few, Gram positive species has been published in a paper on revision of phylogenetic status of *Clostridium bryanti* (Zhao *et al.*, 1990). The dendrogram includes other distantly related Clostridia, *Clostridium pasteurianum, Clostridium barkeri* and *C. sticklandi*, as well as the aerobic actinomyecete *Arthrobacter globiformis, Bacillus subtilis*, and representatives of the third and fourth subdivision *Heliobacterium chlorum* and *Megasphera elsdeni.* To my knowledge the dendrogram has not been used as a base for broader evolutionary speculations, but it seems able to shed light on some important questions. Thus, *Arthrobacter globiformis* clusters with *Bacillus subtilis* , which might be interpreted such that their corresponding sublines shared common aerobic ancestry. Aerobiosis may therefore be a more ancient trait in the phylum than previously thought, because of the great phylogenetic distance between the corresponding species. The aerobic cluster branches off from the clostridia at the base of the tree and it appears that the divergence between the *Arthrobacter globiformis* and *Bacillus subtilis* is approximately of the same magnitude as found between the most distantly related anaerobic clostridia. Aerobiosis could even be the most primitive

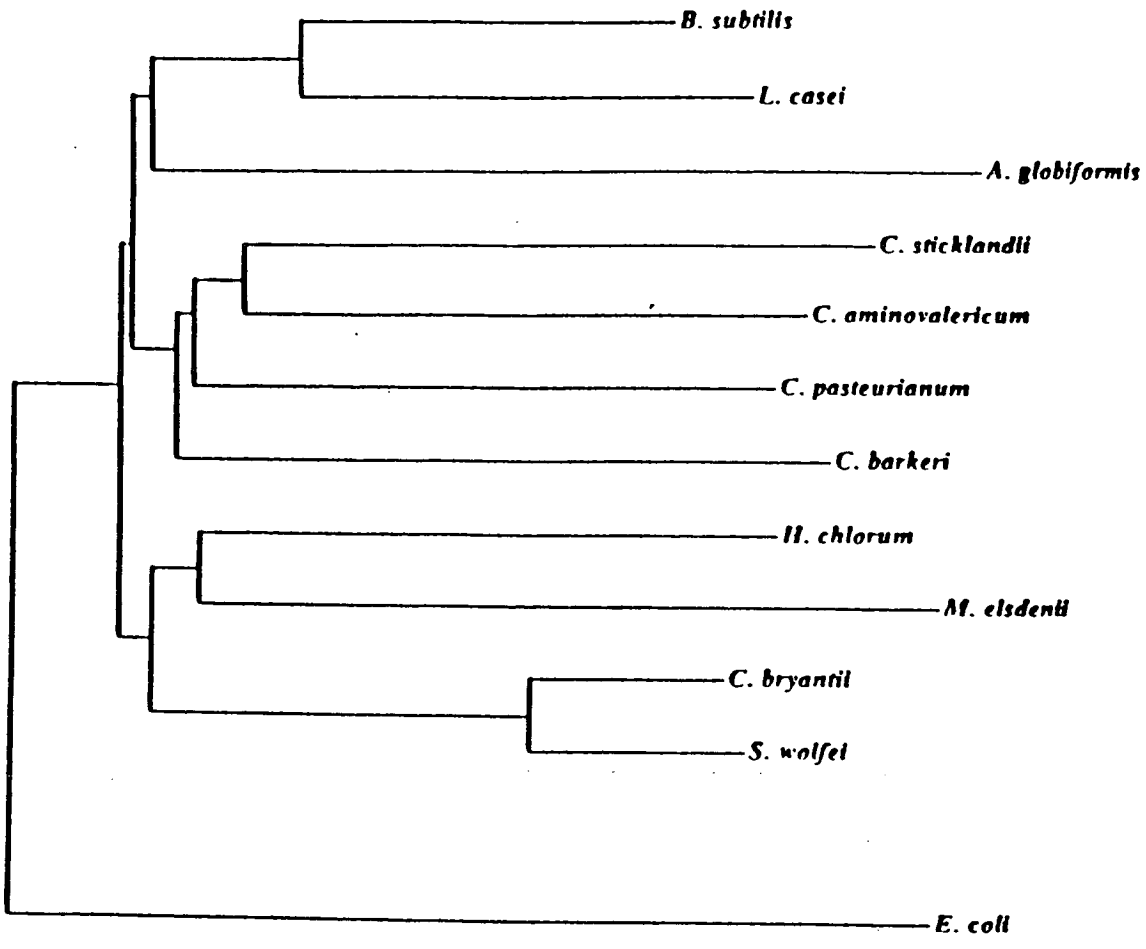


Figure 6.23. Phylogentic reconstruction for a selection of Gram positive species.

The tree is obtained by the augmented distance matrix method (Woese 1987), using the evolutionary distances matrix between complete 16sRNA sequences from the corresponding species and the algorithm of Soete (after Zhao et al., 1990).

metbolic pathway in the phylum (but see below). In the cluster which contains the photosynthetic *Heliobacterium chlorum* deeper branches are found with strictly anaerobic species. This observation does not support the hypothesis for photosynthetic origin of the phylum but it does not preclude it either as distantly related photosynthetic species may still be found within the phylum. Even if they are not found it could be argued that Gram positive bacteria had largely discarded this pathway. An alternative explanation would be lateral transfer which some authors invoke as a possible factor in the evolution and distribution of the photosynthetic machinery (Meyer, 1991).

On the basis of the classification scheme of cytochromes c presented in this thesis it is proposed that the electron transport pathways of aerobiosis in purple bacteria and Gram positive bacteria use basically the same cytochrome c structural types. This, along with the sequence information on other components of the electron transport pathway (section 6.1.2), essentially argues against independent biochemical evolution of aerobic respiration in these different lines of descent. It does not preclude the involvement of transfer.

Interestingly, both phyla contain Class I cytochromes c of the same groups, but the cytochrome c subclass composition of the electron transport chain in the genus *Bacillus* differs markedly from that of any specific subdivision of the purple bacteria, (alpha beta or gamma, section 6.7). In *Bacillus* c2 homologue (the cytochrome c(aa3)) appears to co-exist with so called small bacterial cytochromes, the ID subclass (c7) and even c4 homologues, represented by the BacI subclass cytochromes and the split alpha cytochromes in the genus of *Bacillus.*

The c(aa3) cytochromes from *Bacillus* and *Thermus* are basically the same, both in primary structure and in functional roles, as fused parts of the cytochrome c oxidase. The sequence similarity of other components of the oxidase is also very high (Ishizuka *et al.*, 1990). This conservation of the oxidase is indeed remarkable considering the phylogenetic distance between the corresponding species and argues for a close relationship between their electron transport chains. Furthermore, the split alpha cytochrome in the genus *Bacillus*, which has been partially sequenced from *Bacillus azotoformans* (P3&P4-c555, this work) has a physiological counterpart in *Thermus* (Hon-Nami and Oshima, 1978). It can be argued that the presence of the oxidase in either genus is a consequence of a lateral transfer, as species of the genera, *Thermus*, *Bacillus* and *Clostridium* share hot spring habitats. One possible scenario could involve a lateral

transfer of the genes, coding for the components of the aerobic pathway in an ancestral *Thermus* species, into an ancestral species of anaerobic sporeforming *Clostridium* species. This would perhaps resolve the problem of the enigmatic presence of a fully developed electron transport chain in a phylum with such distinctive anaerobic substructure, as has be noted by different authors and referred to above.

The a-type cytochrome components of the oxidase from *Bacillus* PS3 and *Thermus thermophilus* have clear homologues in such diverse species as *Paracoccus denitrificans* and *Escherichia coli*. Even the organisation of the operon coding for the different genes of the aa3 oxidase subunits is the same in *Escherichia coli* and *Bacillus* PS3 (Ishizuka *et al.*, 1990; Stouthamer, 1992). The *Thermus*/Deinococci group branches from the stem of the eubacterial tree earlier than most other aerobic species. As the oxidase, a terminal transport protein in the aerobic electron transport branch, is also found in such diverse species of later branching phyla it can be argued that aerobiosis is primitive to them all. Consequently, this contradicts the conclusion of Woese (1987), based on the mapping of phenotypes onto the phylogenetic 16S rRNA tree, that aerobiosis has evolved independently in different lines of descent in the eubacterial kingdom and even twice in the phylum of Gram positive bacteria. Excluding extensive interspecies gene exchange, the only other available explanation is that oxygen was available before the divergence of the phyla. Oxygen at such an early stage of eubacterial history appears to contradict geological information, but recently it has been proposed that oxygen was produced much earlier than was previously thought. It is proposed that aerobic respiration started in the early Archean, 3800-2500 Myr. ago (Towe, 1990). Perhaps an early appearance of oxygen could explain the explosive radiation of eubacteria from more or less the same point in the eubacterial tree (figure 6.21a). Certainly the cytochrome c oxidase appears to predate this major divergence. The phylum of green non sulfur bacteria, which is the second deepest eubacterial branching contains both strictly aerobic bacteria, the genera *Herpetosiphon* and *Thermomicrobium*, and obligate anaerobic photosynthetic species of the genus *Chloroflexus*. *Thermatoga*, an earlier branch in the tree, is strictly anaerobic and fermentative in its mode of living (Woese, 1987). Following a similar line of argument, the distribution of the cytochrome c2 variants seems to indicate that c2 cytochrome has been lost from the *Pseudomonas* metabolic lineage. Its respiratory pathway is likely to be an offshoot from

the main aerobic electron transport mode. As such it may not predate the photosynthetic metabolism, but aerobiosis may have in some lineages.

Phylogenetic analysis of 16S rRNA sequences indicates a special relationship between Gram positive bacteria and cyanobacteria (Woese, 1987). Also, Sone and coworkers (Fujiwara et. al., 1993) have emphasised a relatedness of the BacI cytochrome from *Bacillus* PS3 to algal (soluble f) cytochromes, and postulated that the split alpha cytochrome c from the same species is a counterpart to f6 cytochrome in the cyanobacteria. In my opinion this is not supported by sequence analyses of cytochromes c from either phyla. Arguments have been given in the preceding discussion that the BacI sequences are structural homologues to the ID subclass sequences in purple bacteria (see section 6.3.4). Furthermore, the partial sequence of the split alpha cytochrome c in *Bacillus azotoformans* indicates that it is a proper Class I cytochrome, which the f6 cytochrome are not (see section 6.3.3). Interestingly, the f6 cytochromes have counterparts in photsynthetic species belonging to the alpha subgroup of the purple bacteria (Bartsch, 1991). Homologous cytochromes have not been isolated from other bacterial groups. The evidence for its presence in the genera *Chromatium, Pseudomonas, Heliobacterium, Chloroflexus* and *Bacillus* is the presence of Rieske-type iron-sulphur protein, which is detected by an electron paramagnetic resonance signal (EPR), along with the presence of membranous b- and c type cytochromes (Trumpower, 1990). In the light of the sequence information gained in this work on the P3&P4-c555 cytochrome, which has distinct characteristics of a Class I cytochrome, the above cited evidences must be considered tentative. Thus, for the time being at least, the c1/f6 cytochrome distribution appears to be anomalous and such a pattern would best be explained by lateral transfer. The c1 and f6 cytochromes show enough homology to allow alignment (Moore and Pettigrew, 1990), but the low identity between the c1 and the f6 sequences indicates that, if the corresponding genes have been laterally transferred from one to the other genus, it must have occurred far back in the ancestry of the corresponding species. Meyer (1991) furthermore reports that a Class I cytochrome c (an unpublished sequence), from *Ectothiorhodspira halophila* shows affinity to algal (soluble f) cytochromes c (c6). This affinity is also revealed in a dendrogram obtained by multiple alignment analyses of cytochromes c of Class I, by Taylor (1990).

While similar, the basic trunk of the aerobic electron transport chain appears to be constructed differently in *Bacillus* from the one in purple

bacteria. It will therefore be interesting to compare other metabolic branches from these phyla, such as dentrification and methylotrophy, to see if their constructions also turn out to be different. The components of electron transport chain such as cytochromes c may have, at different times, been recruited for similar functions in different species, but the distribution of some cytochromes correlates with the distribution of particular metabolic pathways (Pettigrew and Moore, 1991; Stouthamer, 1992). Thus the ID subclass is found in bacteria or families of bacteria able to denitrify. The denitrifying species *Bacillus azotoformans* contains a cytochrome c (a BacI sequence), which appears to be a structural homologue of the ID subclass cytochromes. It may indicate a common evolutionary origin of the denitrification pathway in these distantly related species.

The occurrence of denitrification in the Gram positive bacteria, is an interesting case for evolutionary study. If one accepts anaerobic photosynthetic ancestry for the Gram positive bacteria, the radiation of the phylum could have been well underway at the onset of aerobic conditions and subsequent emergence of aerobic respiration. Since denitrification is thought to have come in to existence after oxygen accumulated in the atmospere and is dependent on high redox potential cytochromes c (Jones, 1985), it can only mean that the denitrification pathway in Gram positive bacteria has evolved independently from those in other phyla or that the essential components were aquired by lateral transfer. On the other hand, if aerobic conditions were established before the major eubacterial divergence, as can be argued (see above) the distribution of the denitrification metabolism may be explained by 'vertical' transmission of the essential components from the common ancestral species to all denitrifiers.

The dendrogram in figure 6.23, is also interesting for speculations on the evolution of the Gram positive cell wall. On the basis of the Gram negative characteristics of *Megaspheara elsdanii, Sporamusa, Selenomonas* and *Heliobacterium chlorum.*, Stackebrandt (1985) has proposed that the ancestral cell wall type of Gram positive bacteria was likely to have been Gram negative. These species cluster together in the dendrogram in figure 6.23. However, this cluster includes one species with a typical Gram positive cell wall, *Syntophospora bryantii.* Because of the topology of this dendrogram, the most likely conclusion is that the Gram positive ultrastructure of the wall is ancestral to the phylum, but it might be more ancient still. The genus *Thermus* appears to have a Gram negative cell wall

structure and the mode of cell division is by invagination rather than septum formation in the manner of Gram positive bacteria. However, *Thermus* has ornithine peptidoglycan in the cell wall, which has not been reported in Gram negative bacteria and also branched fatty acids typical of Gram positive bacteria. Furthermore, it uses menaquinone and is very sensitive to penicillin, actinomycin D and novobicin that is more characteristic of Gram positive bacteria than of Gram negative bacteria. (Williams, 1992). On the basis of these Gram positive features it has been postulated that the genus predates the Gram positive/Gram negative dichotomy. Indeed this is supported by16S rRNA analyses by Woese and co-workers who place the genus *Thermus* in a earlier brancing eubacterial phylum with the *Deinococcus* (Weisburg *et al.*, 1989). The *Deinococcus* are the only bacteria outside the 'Gram positive phylum proper' to have a positive response to the Gram stain. Furthermore it has been shown to have ornithine peptidoglycan in its cell wall. It is strictly aerobic and catalase positive and has been shown to contain cytochromes (Murray, 1986). Contradicting this is the presence in *Deinococcus* of some typical Gram negative features such as an outer membrane, and it also has a typical Gram negative fatty acid profile. These genera are therefore neither truly Gram positive nor Gram negative, and consequently the phylum may represent the extant species of the kind of bacteria, which were ancestral to both. On a more speculative note, however, it can be postulated on the basis of the remarkable similarity of the cytochrome c oxidase of *Bacillus* PS3 to its counterpart in *Thermus thermophilus*, as noted above, in conjunction with above evidence for a special affinity of *Thermus* to Gram positive bacteria, that the genus of *Thermus* is essentially Gram positve and the ancient features of the aerobic pathway have been conserved in the Gram positve phylum.

6.8. CYTOCHROME C SEQUENCES AND THE PHYLOGENY IN THE GENUS *BACILLUS*

Three complete haem domain sequences of the same structural class have been reported from as many species of the genus of *Bacillus*, *B. licheniformis*, *Bacillus* PS3, and *B. subtilis*. The fourth complete haem domain sequence is reported in this thesis from *B. azotoformans*. A partial sequence has been reported from *B. halodenitrificans* and another sequence of the same type has been partially sequenced from *B. licheniformis*. Furthermore, physicochemical evidence indicates that there is a second sequence of this type in *B. azotoformans* .

In figure 6.24 is shown the dendrogram obtained by the UPGMA clustering method from alignment of the four complete haem domain sequences from the genus *Bacillus*. As there are two cytochromes of the same structural subclass in some, if not all, *Bacillus* species it makes, however, any cladistic interpretation of the sequence information highly doubtful. The prerequisite for phylogenetic analysis based on sequence data is, that the sequences should be homologous. The great similarity between the *B. licheniformis* c551 and the *B. subtilis* c550 cytochromes (74%), should be noted, which indicates that they are homologous. Also, a marked similarity is found between the cytochrome c550 partial sequence from *Bacillus halodenitrificans*, including the N-terminal membrane anchorage part, to the c551 cytochrome sequence of *Bacillus* PS3. According to the corresponding descriptions in Bergey´s manual there are few phenotypic characters that distinguish between *Bacillus licheniformis* and *Bacillus subtilis* and their close relationship is supported by various phenetic studies, numerical taxonomic studies (Priest, 1981), lipid analysis (Minnikin,1981), Api tests (Logan, 1981), and pyrolysis studies (O´Donnell, 1981).

Another point worth noting is, that the sequences, apart from the two sequence pairs, the *B. licheniformis* and *B. subtilis* cytochromes, and the *Bacillus* PS3 and *B. halodenitrificans*, appear to be approximately equidistant from each other at the 40% level. This is for example observed between *B. azotoformans* or the *Bacillus* P3 sequences and either of the isologous *Bacillus subtilis* and *B. licheniformis* cytochromes. This may indicate that some of pairwise distances reflect the fact that the corresponding sequences have reached mutational equilibrium relative to the timespan since their divergence from the ancestral sequence. This duplication appears to have happened in the genus before the separation

of *Bacillus licheniformis* and *Bacillus azotoformans*. However it should be mentioned, that while there is sequence evidence for the presence of this duplication in *B. licheniformis*, there is only physicochemical evidence for it in *B. azotoformans* (this work). The mutation equilibrium may best be explained by constraints acting on the structure (Meyer *et al.*, 1986).

Interestingly from a taxonomic point of view, the divergence of the BacI cytochromes between species from the genus is much greater than the divergence between the proposed homologous ID subclass sequences between different genera. The minimum identity in the former genus is 38%, but in the latter it is 59% (figure 6.5b). Either this indicates that; selection constraints are far greater on ID subclass cytochromes because there is only one of its kind per species compared

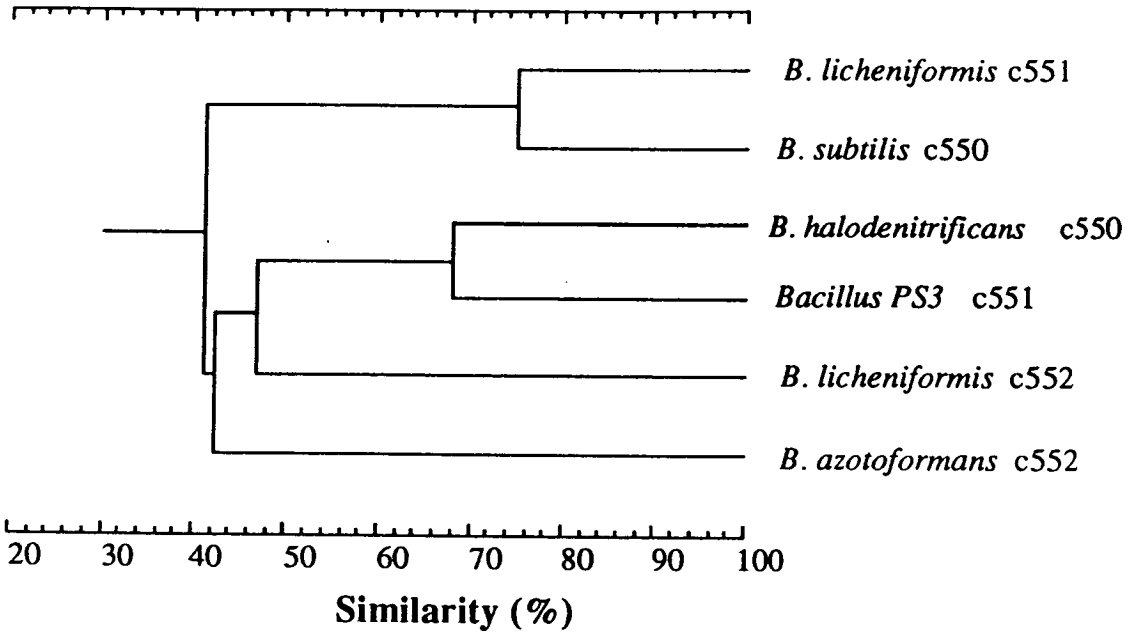


Figure. 6.24. An UPGMA (%) similarity dendrogram, cast from the % identity values for the Bacillus cytochromes c in the matrix in figure 6.5b.

to two in *Bacillus* species, or; it indicates that the genus of *Bacillus* is older. This latter perhaps reflects the great weight of relatively few traits in the taxonomic specification for the genus, a positive Gram stain response, aerobiosis and sporeforming, instead of greater emphasis on metabolic or physiological characteristics.

APPENDICES

APPENDIX A

Results of autamated sequencing

The following tables show the yields of amino acids of each sequencing cycle. The highest value is given and all values 10% or larger of the highest value in each cycle. The figures show the chromograms of two cycles in two the sequencing of two different peptides. Figure A1 show the proposed peak of methionine sulphone and its retention time. Figure A2 shows the peak and retention time of proposed modified tryptophan.

Table A.1. P1-c552 N-terminus sequencing. Results of the automated sequencing.

Cycle	Amino acid	A Ala	R Arg	N Asn	D Asp	C Cys	E Glu	Q Gln	G Gly	H His	I Ile	L Leu	K Lys	M Met	F Phe	P Pro	S Ser	T Thr	W Trp	Y Tyr	V Val
1	G				22				342		21										
2	Q						108	406													
3	A	515																			
4	T																	308			
5	E						314														
6	Q						82	356													
7	A	461																			
8	A	377																			
9	P															357					
10	A	442																			
11	A	360																			
12	D				200																
13	D				250																
14	P				60											215					
15	E				30		159									23					
16	G								166							23					
17	M						29							166							
18	V																				167
19	K			14					14				136	17							
20	A	124																			14
21	S		4	7													39				
22																					
23	K										11	12	13	10		8			7		11
24	A	30	3	7			3		5												
25	S	5		2	16		3		3								20				
26	D				20				5	3							6		2		
27	G				19				25												
28	D				21			15	10												

The values are the obtained yields (pmol) of amino acids in each cycle of the AB477A sequencer run.

Table A.2. P1-c552. Results of the automated sequencing of peptide T1

Cycle	Amino acid	A Ala	R Arg	N Asn	D Asp	C Cys	E Glu	Q Gln	G Gly	H His	I Ile	L Leu	K Lys	M Met	F Phe	P Pro	S Ser	T Thr	W Trp	Y Tyr	V Val
1	A	1738																			
2	S																566				
3				9			2	4	9		1								2		
4	A	1464																			
5	S																374				
6		11						1	3				2								
7	H									174											
8	G								703												
9	Q						132	668													
10	N			599																	
11	L											857									
12	E						317														
13	G								515												
14	G								479												
15	V																				549
16	G								425												
17	P															462					
18	A	518																			
19	L											502									
20	A	432																			
21	D				5																
22	D	75			194																
23	V																				296
24	G								235												
25	S		13						10								72				
26	R		56																		
27	R		22																		

The values are the obtained yields (pmol) of amino acids in each cycle of the AB477A sequencer run.

Table A.3. P1-c552. Results of the automated sequencing of peptide T2

Cycle	Amino acid	A Ala	R Arg	N Asn	D Asp	C Cys	E Glu	Q Gln	G Gly	H His	I Ile	L Leu	K Lys	M Met	F Phe	P Pro	S Ser	T Thr	W Trp	Y Tyr	V Val
1	N			841	95																
2	A	739																			
3			7																		
4	P											5				16					3
5	A	687														675					
6	G																				
7									637												
8	V																				15
9	D				320																638
10	P																				
11	A	500														494					
12	K																				
13	A	430											401								
14	E		98				158														
15	V																				
16	I																				405
17	A	276									375										

The values are the obtained yields (pmol) of amino acids in each cycle of the AB477A sequencer run.

Table A.4. P-c552. Results of the automated sequencing of peptide T5

Cycle	Amino acid	A Ala	R Arg	N Asn	D Asp	C Cys	E Glu	Q Gln	G Gly	H His	I Ile	L Leu	K Lys	M Met	F Phe	P Pro	S Ser	T Thr	W Trp	Y Tyr	V Val
1	L																				
2	S											2513									
3	A	1954															850				
4	E																				
5	E						677														
6	I						836														
7	A	1644									1490										
8	N			948		98															
9	I																				
10	I										1360										
11	N			941		94					1362										
12	N			867		85															
13	G			15																	
14	K								768												
15													357								

The values are the obtained yields (pmol) of amino acids in each cycle of the AB477A sequencer run.

Table A.5. P-c552. Results of the automated sequencing of peptide V5·

Cycle	Amino acid	A Ala	R Arg	N Asn	D Asp	C Cys	E Glu	Q Gln	G Gly	H His	I Ile	L Leu	K Lys	M Met	F Phe	P Pro	S Ser	T Thr	W Trp	Y Tyr	V Val
1	V																				
2	I																				751
3	A	718									738										
4	A	708																			
5	W*	11							3							4				23	
6	L											427									
7	A	492																			
8	E						267														
9	Q						69	287													

The values are the obtained yields (pmol) of amino acids in each cycle of the AB477A sequencer run.

Table A.6. P3&P4-c555. N-terminus sequencing. The results of the automated sequencing.

Cycle	Amino acid	A Ala	R Arg	N Asn	D Asp	C Cys	E Glu	Q Gln	G Gly	H His	I Ile	L Leu	K Lys	M Met	F Phe	P Pro	S Ser	T Thr	W Trp	Y Tyr	V Val
1	Q																				
2	T							928													
3	V		(3035)					71													
4	E																	693			
5	F						661														1548
6	D		236													104					
7	K				575										1252						
8	K		570									58				82					
9	D		(8667)										2802								
10	P	78			431		152		205			105	2347								
11	G								181							186	64				
12	Y								924							985					
13	K						107		167			147				144					
14	A	912	10609?						214				1884			166				863	
15	L?													4163?							
16	E?																				
17	A	493	4743?																		
18	S?												2620?								
19	G																173				
20	L?		69	42			213		705			388				122			71		
							199		358		43	388		43		108			68		

The values are the obtained yields (pmol) of amino acids in each cycle of the AB477A sequencer run.

Table A.7. P3&P4-c555. Results of the automated sequencing of peptide T1.

Cycle	Amino acid	A Ala	R Arg	N Asn	D Asp	C Cys	E Glu	Q Gln	G Gly	H His	I Ile	L Leu	K Lys	M Met	F Phe	P Pro	S Ser	T Thr	W Trp	Y Tyr	V Val
1	K																				
2	Y												908								
3	D				445															728	
4	P																				
5	L															869					
6	A	893										877				93					
7	E											89									
8	V						448														
9	Q																				765
10	T						127	591													
11	D																	366			39
12	K				312																
13	V												317								
14	N			231	34																314
15	A	221																			
16	E																				
17	I						87														
18	P										118										
19	S															107					
20	S																32				
21	F															3	27				
22	A	38													37						
23	G																				
24	V								26												
25	V?																				27
																					24

The values are the obtained yields (pmol) of amino acids in each cycle of the AB477A sequencer run.

Table A.8. P3 & P4-c555. Results of automated sequencing of peptide T2

Cycle	Amino acid	A Ala	R Arg	N Asn	D Asp	C Cys	E Glu	Q Gln	G Gly	H His	I Ile	L Leu	K Lys	M Met	F Phe	P Pro	S Ser	T Thr	W Trp	Y Tyr	V Val
1	A	278										61						68			
2	L											387									
3	E	54					206														
4	A	296																			
5	S								53								119				
6	G								192												
7	C*											35									
8	L			25								158									
9	N			104																	
10	C*	2								1		1				1			1		
11	H	6		3					19	11						3					
12	G								61									7			
13	T				5													33			
14	D				32							8									
15	L											40						6	5		
16	T								4									20			
17	G								30												
18	G								17							2					
19	P															19					
20	A	21																	5		
21	A	8							1							1					
22	P															10					
23	A	6										1				1			1		
24	L										1	5						2			
25	T																	3			

The values are the obtained yields (pmol) of amino acids in each cycle of the AB477A sequencer run.

Table A.9. P3 & P4-c555. Results of automated sequencing of peptide T4

Cycle	Amino acid	A Ala	R Arg	N Asn	D Asp	C Cys	E Glu	Q Gln	G Gly	H His	I Ile	L Leu	K Lys	M Met	F Phe	P Pro	S Ser	T Thr	W Trp	Y Tyr	V Val
1	A	3872																			
2	L											3591									
3	S									185							1037				
4	E						1361														
5	F														2053						
6	I										2015										
7	H									321											
8	G								1197												
9	L											1223									
10	S									42							239				
11	A	632																			
12	K												270								

The values are the obtained yields (pmol) of amino acids in each cycle of the AB477A sequencer run.

Table A.10. P3 & P4-c555. Results of automated sequencing of peptide T9

Cycle	Amino acid	A Ala	R Arg	N Asn	D Asp	C Cys	E Glu	Q Gln	G Gly	H His	I Ile	L Leu	K Lys	M Met	F Phe	P Pro	S Ser	T Thr	W Trp	Y Tyr	V Val
1	A	428																			
2	T																				
3	M																	152			
4	P													135							
5	P															97					
6	G															53					
7	M								41							6					
8	F													25							
															22				3		

The values are the obtained yields (pmol) of amino acids in each cycle of the AB477A sequencer run.

Table A.11. P2-c551. N-terminus sequencing. Results of the automated sequencing.

Cycle	Amino acid	A Ala	R Arg	N Asn	D Asp	C Cys	E Glu	Q Gln	G Gly	H His	I Ile	L Leu	K Lys	M Met	F Phe	P Pro	S Ser	T Thr	W Trp	Y Tyr	V Val
1	E	174					483		183								237				
2	P															719					
3	A	617																			
4	D				310																
5	N			333	54																
6	N			303	33																
7	A	363																			
8	G								273												
9	D			4	28											7			3		
10	G						3	5	17		2								4		

The values are the obtained yields (pmol) of amino acids in each cycle of the AB477A sequencer run.

Table A.12. P5-c552. N-terminus sequencing. Results of automated sequencing.

Cycle	Amino acid	A Ala	R Arg	N Asn	D Asp	C Cys	E Glu	Q Gln	G Gly	H His	I Ile	L Leu	K Lys	M Met	F Phe	P Pro	S Ser	T Thr	W Trp	Y Tyr	V Val
1	E	113					315		41												
2	Q						141	475													
3	T		197					45										381			
4	T		208															378			
5	E						316														
6	T		192															319			
7	A	511																			
8	Q						79	396													
9	P															476					
10	N			297																	
11	L											375									
12	A	354																			
13	E						175														
14	H*																				
15	T		78															167			
16	D				155																
17	N			181																	
18	I										230										
19	L											195									
20	V																				209
21	Q						31	153													
22	K												131								
23	G								126												
24	C*																				
25	I										123										
26	A	128																			
27	C*																				
28	H*																				
29	A	79	8																		
30	V	30	11									8									84
31	S		7		5		3									3	21	2			4
32	P															39					
33	I				6		3				27					24	3	3			
34	G		18						20		12										
35	A	30	11						7												
36	K	3	12								3		23						8		
37	G								23				6								

The values are the obtained yields (pmol) of amino acids in each cycle of the AB477A sequencer run.

Figure A1

= Applied Biosystems 477A Protein Sequencer Chromatogram Report =

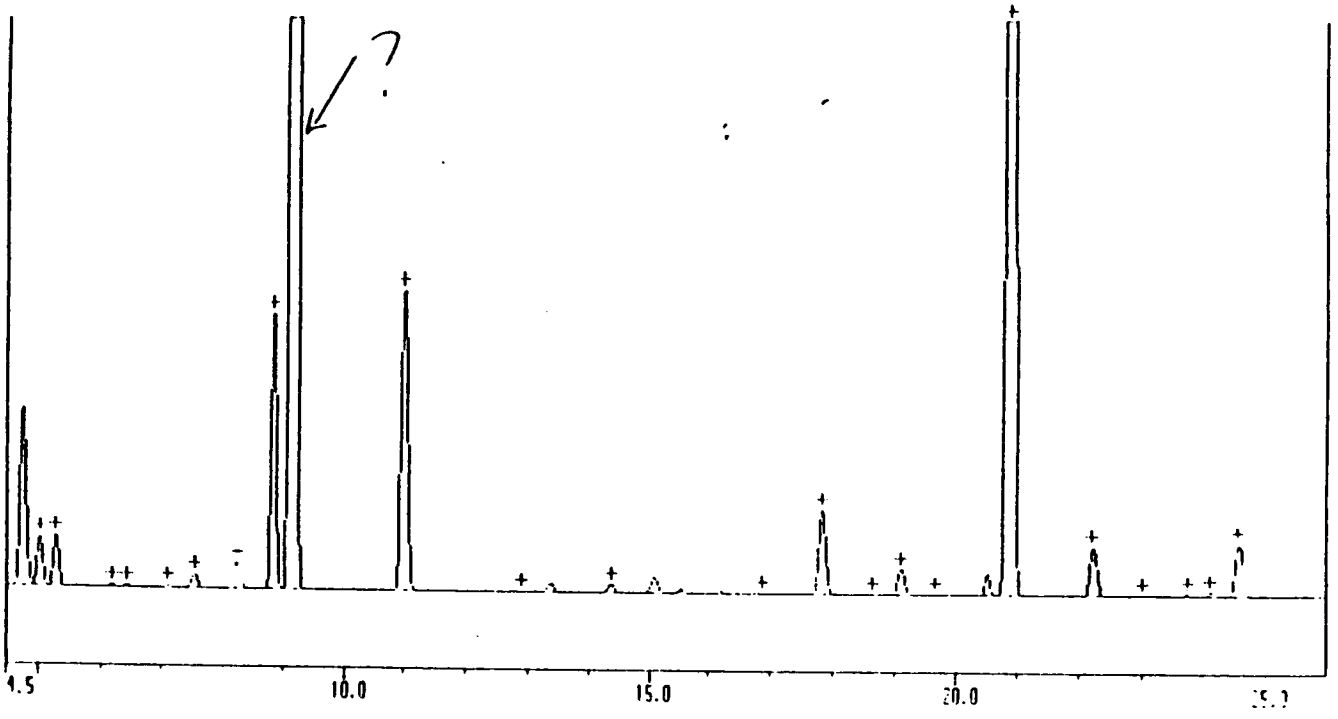
SAMPLE : B.ACIDOFORMANS
[Initiated 1 Jun 1990 12:18pm]

CYCLE SUMMARY :

Reaction cycle	: NORMAL-1	Data collect time	: 0.0 to 29.0 min
Conversion cycle	: NORMAL-1	Data interval	: 1.0 sec
Gradient	: NORMAL-1	Inject volume	: 50 of 132 uL

AMINO ACID # 3 [1 Jun 1990 3:37pm] 0.0:20 FU

Filtered Data



Retention Time: Minutes

PEAK TABULATION : (100% injection) Calibration : LINEAR

Peak ID	R.Time (min)	C.Time (min)	Height (uAU)	Pmol	Peak ID	R.Time (min)	C.Time (min)	Height (uAU)	Pmol
	4.78		8211		PRO	17.95	17.97	3876	20.74
ASP	5.05	5.15	2331	4.44	MET	18.55	18.87	12	0.04
ASN	5.30	5.40	2361	5.59	VAL	19.13	19.27	1219	4.71
SER	6.20	6.30	114	0.71	PEC	19.68	20.02	49	0.36
GLN	6.42	6.53	111	0.43		20.52		1017	
THR	7.12	7.25	93	0.35	DPT	20.87	21.00	133737	779.72
GLY	7.57	7.68	645	2.38	TRP	22.23	22.29	2292	5.35
GLU	8.23	8.40	1218	3.19	PHE	23.03	23.13	54	0.20
DMP	8.97	8.98	12657	220.20	ILE	23.75	23.92	81	0.41
	9.18		175074		LYS	24.13	24.25	195	0.56
ALA	11.02	11.15	13770	60.30	LEU	24.53	24.72	2484	12.03
HIS	12.90	12.95	3	0.03					
TYR	14.38	14.58	390	1.40					
	15.10		708						
ARG	16.88	17.00	57	13.44					

Tabulation threshold : 500 uAU

Figure A2

= Applied Biosystems 477A Protein Sequencer Chromatogram Report =

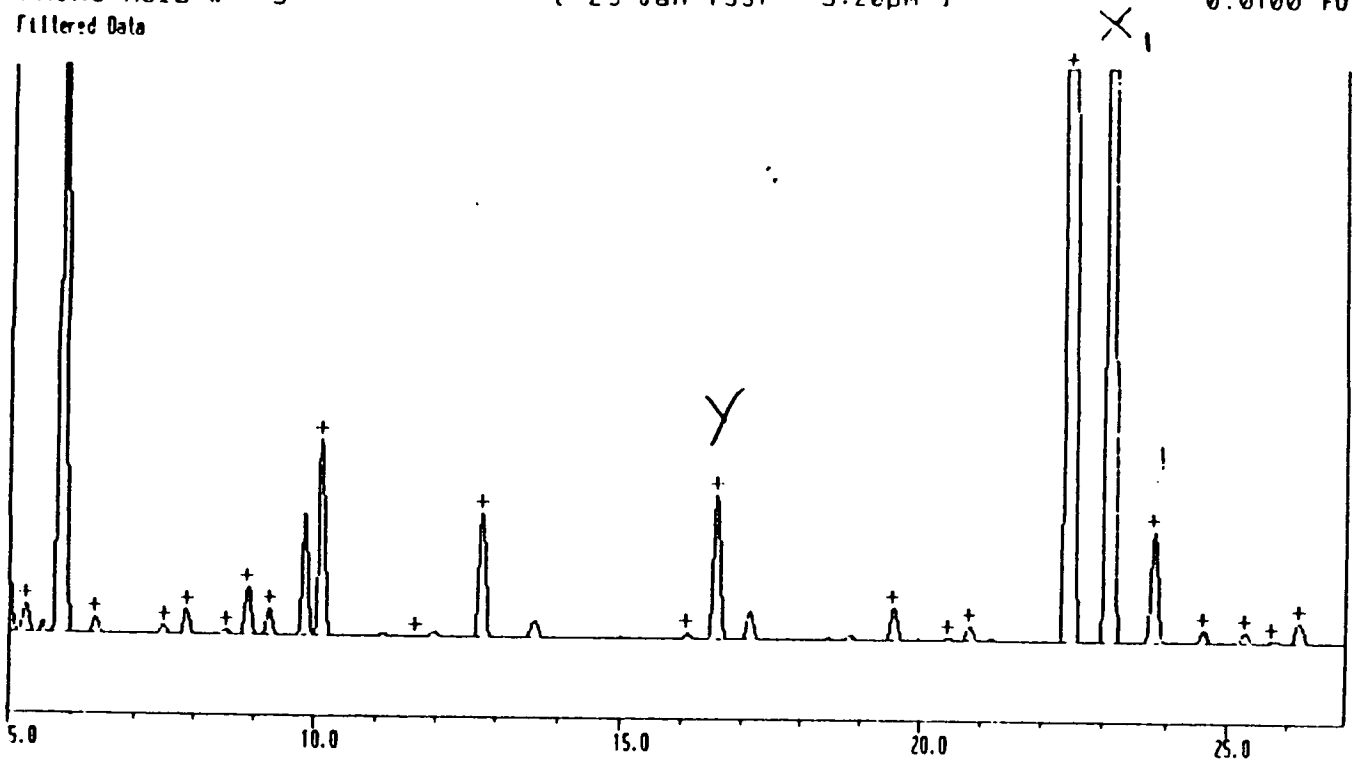
SAMPLE : IIIC GUDMUNDER
[Initiated 29 Jan 1991 12:31pm]

CYCLE SUMMARY :

Reaction cycle	: NORMAL-1	Data collect time	: 0.0 to 29.0 min
Conversion cycle	: NORMAL-1	Data interval	: 1.0 sec
Gradient	: NORMAL-1	Inject volume	: 50 of 150 uL

AMINO ACID # 5 [29 Jan 1991 5:20pm] 0.0100 FU

Filtered Data



Retention Time: Minutes

PEAK TABULATION : (100% injection) Calibration : SDP191

Peak ID	R.Time (min)	C.Time (min)	Height (uAU)	Pmol	Peak ID	R.Time (min)	C.Time (min)	Height (uAU)	Pmol
ASP	5.27	5.35	1248	2.66	PRO	19.53	19.63	1488	7.04
	5.85		31359		MET	20.47	20.53	135	0.51
ASN	6.37	6.47	738	2.32	VAL	20.80	20.90	684	2.34
SER	7.48	7.60	393	2.60	DPT	22.45	22.50	210084	1084.70
GLN	7.87	8.00	1128	5.04		23.12		84915	
THR	8.53	8.62	228	1.23	TRP	23.80	23.78	5091	14.47
GLY	8.88	9.00	2193	11.96	PHE	24.62	24.65	573	2.38
GLU	9.25	9.35	1206	3.84	ILE	25.30	25.32	516	3.02
	9.83		5595		LYS	25.73	25.77	162	0.60
DMP	10.12	10.22	9081	155.60	LEU	26.18	26.23	945	5.19
HIS	11.70	11.83	36	0.39					
ALA	12.78	12.90	5763	28.21					
ARG	18.10	16.10	267	3.62					
TYR	16.60	16.60	6597	24.85					
	17.13		1293						

Tabulation threshold : 999 uAU

APPENDIX B

Gribskov (1987) has devised a method for searching protein databases for distantly related proteins or specific sequence motifs. The method uses a position-specific scoring matrix (profile), which expresses the information in a group of homologous sequences previously aligned by structural or sequence similarity. This means that each amino acid gets a specific score value in the different positions of the protein sequence. The amino acid score at a particular position is derived from the amino acid composition at this position in the prealigned sequences and is calculated from some basic Dayhoff relatedness odd logs matrix, such as the PAM 100 matrix or any other scoring matrix which can for example be based on some quantifiable physicochemical property of amino acids (e.g. hydrophobicity). Accordingly the same amino acid scores differently in different positions. The score reflects the likelihood for any amino acid to be found in a particular position in the protein structure. A subclass or any sequence group for that matter, can be reduced into such a numerical group-specific profile matrix where some sequence positions are enhanced in importance and others devalued. Similarity should therefore have a greater chance of being significantly detected. This requires that a number of homologous sequences has to be available, which limits the usefulness of the method. The outline of the method is presented in figure B.1.It shows a part of a profile obtained from the corresponding alignment of the BacI cytochromes c. The position specific score of a any amino acid is the average Dayhoff score (PAM 100, 20 x 20 pairwise scoring matrix for amino acids, Dayhoff, 1978)) of that amino acid to every amino acid in that particular position of the alignment.

In this discussion the profile method has been adapted for a graphical representation of sequence conservation to help clarify and analyse the primary structures of different cytochrome c sequence groups. It summarises and makes the information in the alignments more accessible for interpretation and enables a visual comparison of sequence characteristics between different subclasses of proteins of the same family, or between sequence motifs in apparently unrelated protein families.

The profile scoring matrix obtained from a group of prealigned protein sequences is transformed in the following way: Each column of the profile matrix is treated individually. The mean amino acid score for each column

Ala	3	2	2	0	0	-3	-5	-4	2	0	0	-5	2	2	-5	-5	1	-2	-1	PROFILE
Cys	-7	-7	-6	-11	-10	-7	-5	-11	-8	0	-6	14	-6	-6	14	-6	-8	-11	-10	
Asp	1	4	-3	5	3	-7	-9	-1	0	0	3	-11	-3	-1	-11	-1	-1	1	6	
Glu	1	2	-2	8	5	-5	-8	-1	2	0	0	-11	-3	0	-11	-2	-2	4	3	
Phe	-8	-9	-8	-11	-10	-1	5	-11	-9	0	-6	-10	-5	-8	-10	-4	-8	-11	-9	
Gly	-1	0	1	-2	0	-7	-9	-5	-2	0	-1	-8	0	1	-8	-7	8	-4	-1	
His	-4	-3	-4	-2	-2	-7	-3	-1	-1	0	1	-6	-6	-3	-6	11	-7	1	1	
Ile	-3	-5	-6	-5	-6	7	-1	-4	-4	0	-4	-5	-1	-5	-5	-7	-7	-5	-5	
Lys	-3	-3	-4	-2	-3	-3	-9	6	-3	0	0	-11	-4	-3	-11	-3	-5	1	-1	
Leu	-6	-8	-6	-7	-6	3	-2	-5	-4	0	-6	-12	-4	-6	-12	-5	-8	-5	-8	
Met	-4	-6	-6	-6	-6	5	-4	0	-3	0	-5	-11	-3	-4	-11	-7	-8	-2	-7	
Asn	1	2	-2	1	0	-4	-4	1	-1	0	6	-8	-2	0	-8	2	-1	0	6	
Pro	-1	-2	5	-3	-3	-6	-9	-3	0	0	-2	-6	-2	-1	-6	-2	-3	-2	-4	
Gln	-2	-1	-2	4	3	-4	-8	2	4	0	-2	-11	-4	0	-11	4	-5	5	0	
Arg	-5	-5	-4	-5	-4	-4	-8	3	-2	0	-3	-6	-5	-3	-6	1	-8	0	-5	
Ser	1	2	1	-2	-2	-4	-5	-2	-1	0	3	-1	0	2	-1	-4	1	-3	1	
Thr	2	0	-1	-3	-3	-1	-5	-2	-1	0	1	-5	1	-1	-5	-5	-3	-3	-1	
Val	-2	-4	-3	-5	-5	4	-2	-6	-3	0	-5	-4	0	-3	-4	-6	-4	-5	-6	
Tyr	-11	-10	-12	-14	-13	-12	7	-10	-11	0	-7	-13	-12	-10	-13	-7	-13	-11	-11	
Trp	-7	-8	-10	-7	-9	-5	-5	-10	-8	0	-4	-2	-7	-8	-2	-1	-11	-9	-6	
	A	D	P	E	E	I	Y	K	Q	-	N	C	A	A/S	C	H	G	Q	D	CONCENSUS SEQUENCE
score	3	4	5	8	5	7	7	6	4	0	6	14	2	2	14	11	8	5	6	115 =Cumulative consensus score
	A	A	A	E	Q	I	F	K	Q	-	N	C	A	S	C	H	G	Q	D	ALIGNMENT OF Bacl
	D	D	P	E	G	M	V	K	A	-	S	C	G	Q	C	H	G	Q	N	CYTOCHROMES C
	A	S	P	E	E	I	Y	K	A	-	N	C	I	A	C	H	G	E	N	
	T	D	G	E	E	I	Y	Q	Q	-	N	C	T	G	C	H	G	K	D	

PAM 100 Dayhoff amaino acid scoring matrix.

	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Tyr	Trp
Ala	6	-5	-1	0	-7	1	-5	-3	-4	-5	-3	-1	1	-2	-5	2	2	0	-11	-6
Cys	-5	14	-11	-11	-10	-8	-6	-5	-11	-12	-11	-8	-6	-11	-6	-1	-5	-4	-13	-2
Asp	-1	-11	8	5	-11	-1	-1	-6	-2	-9	-8	4	-4	1	-6	-1	-2	-6	-13	-9
Glu	0	-11	5	8	-11	-2	-2	-5	-2	-7	-6	1	-3	4	-5	-2	-3	-5	-14	-7
Phe	-7	-10	-11	-11	12	-8	-4	0	-11	0	-2	-6	-9	-10	-7	-5	-6	-5	-2	6
Gly	1	-8	-1	-2	-8	8	-7	-7	-5	-8	-8	-1	-3	-5	-8	1	-3	-4	-13	-11
His	-5	-6	-1	-2	-4	-7	11	-7	-3	-5	-7	2	-2	4	1	-4	-5	-6	-7	-1
Ile	-3	-5	-6	-5	0	-7	-7	9	-4	2	2	-4	-6	-5	-4	-4	-1	5	-12	-4
Lys	-4	-11	-2	-2	-11	-5	-3	-4	8	-6	1	1	-4	-1	3	-2	-1	-6	-9	-10
Leu	-5	-12	-9	-7	0	-8	-5	2	-6	9	4	-6	-5	-3	-7	-7	-5	1	-7	-5
Met	-3	-11	-8	-6	-2	-8	-7	2	1	4	13	-5	-6	-2	-2	-4	-2	1	-11	-8
Asn	-1	-8	4	1	-6	-1	2	-4	1	-6	-5	7	-3	-1	-3	2	0	-5	-8	-3
Pro	1	-6	-4	-3	-9	-3	-2	-6	-4	-5	-6	-3	10	-1	-2	1	-1	-4	-11	-11
Gln	-2	-11	1	4	-10	-5	4	-5	-1	-3	-2	-1	-1	9	1	-3	-3	-5	-11	-9
Arg	-5	-6	-6	-5	-7	-8	1	-4	3	-7	-2	-3	-2	1	10	-1	-4	-6	1	-10
Ser	2	-1	-1	-2	-5	1	-4	-4	-2	-7	-4	2	1	-3	-1	6	2	-4	-4	-6
Thr	2	-5	-2	-3	-6	-3	-5	-1	-1	-5	-2	0	-1	-3	-4	2	7	-1	-10	-6
Val	0	-4	-6	-5	-5	-4	-6	5	-6	1	1	-5	-4	-5	-6	-4	-1	8	-14	-6
Tyr	-11	-13	-13	-14	-2	-13	-7	-12	-9	-7	-11	-8	-11	-11	1	-4	-10	-14	19	-2
Trp	-6	-2	-9	-7	6	-11	-1	-4	-10	-5	-8	-3	-11	-9	-10	-6	-6	-6	-2	13

Calculation of the score for alanine in position 1 (boxed gray in the profile matrix)) of the alignment.

is calculated by taking into account the observed amino acid frequencies in proteins sequenced to date (each column is considered as a slot independent from all others, where amino acids fall into, randomly, and score). This value is subtracted from each score value in the in the corresponding column and a new derived profile scoring matrix obtained, which gives the sore as Dayhoff Pam100 units above the average score in each position.

The consensus scores of the alignment (the score value of the amino acid which has the highest scoring value in each column) are then plotted against their sequence position If an invariant residue is found in the same position of two different protein groups, they will get the same numerical score and the same coordinates. Single sequences or consensus sequences of other subclasses which have been aligned with the sequence alignment from which the profile is derived, can be scored in the profile matrix and plotted on the same graph for visual inspection and comparison. One, two and three standard deviations are calculated from the mean score for the whole derived profile, again by taking into account the observed amino acid frequencies in proteins. These values are plotted as reference lines, onto the graph. The profile consensus sequences can the used as parameters of similarity between sequence classes, and pairwise cumulative scores of consensus sequences calculated from the profile matrices.

A rudimentary affinity score between two subclasses, subclassI and subclassII, is calcuated thus:

Affinity score between subclassI and subclassII =

$$((S.Cons.I_{ProfileII} + S.ConsII_{ProfileI.})/2)/ ((S.ConsI_{ProfileI} \times (L_{(IxII)}/L_{ConsI}) + (S.ConsII_{ProfileII} \times (L_{(IxII)}/L_{ConsII}))/2$$

$S.ConsI_{ProfileII}$ = The cumulative score of the consensus sequence for subclassI in the profile of subclassII

$S.ConsII_{ProfileI}$ = The cumulative score of the consensus sequence for subclassII in the profile of subclassI

$S.Cons.I_{ProfileI}$ = The cumulative score of the consensus sequence for subclassI in its own the profile (that of subclassI)

$S.ConsII_{ProfileII}$ = The cumulative score of the consensus sequence for subclassII in its own profile (that of subclassII)

$(L_{(IxII)}$ = The number of positions the consensus sequences of the two subclasses have in common.

L_{ConsI} = The number of positions in consensus sequence of subclassI

L_{ConsII} = The number of positions in consensus sequence of subclassII

APPENDIX C

Biochimica et Biophysica Acta, 1058 (1991) 52–55

ADONIS 0005272891001281

BBABIO 43404

Two structurally different cytochromes *c* from *Bacillus azotoformans*: on the evolution of Gram-positive bacteria

Gudmundur Oli Hreggvidsson

Institute of Cell and Molecular Biology, Division of Biological Sciences, University of Edinburgh, Edinburgh (U.K.)

(Received 5 February 1991)

Key words: Cytochrome c-552; Cytochrome c-555; Sequence; Evolution; (*B. azotoformans*)

c-552 and split-alpha c-555 cytochromes from *Bacillus azotoformans* are classified on the basis of partial sequence information. The haem-containing polypeptides are postulated to be structurally equivalent to small IC and ID subclass cytochromes found in purple bacteria.

Introduction

Bacillus azotoformans (ATCC 29788 (DSM 1046)) is a Gram-positive soil bacterium capable of denitrifying as well as aerobic respiration [1]. It produces an abundance of *c*-type cytochromes, several of which have been isolated and are currently under study in laboratory. In this paper are presented partial sequences of two *c*-type cytochromes from *B. azotoformans* with distinctly different spectral and structural characteristics; one a *c*-552 cytochrome, the other a cytochrome *c*-555 with a split alpha band [2]. Their structural characteristics and evolutionary significance are discussed in the context of other relevant sequence information from the phylum of Gram-positive bacteria.

The aerobic respiration chain in the genus *Bacillus* is similar in general characteristics to its counterpart in Gram-negative bacteria. Components of the same or similar biochemical and biophysical characteristics are found in both pathways. Thus *c*-type and *b*-type cytochromes as well as *a*- and *o*-type cytochrome oxidases have been reported from *Bacillus* species [3–7]. However, the evolutionary significance of the observed similarities has still to be assessed and this is feasible only on the basis of primary sequence information. Relationships ascertained by the sequencing of electron transport proteins provide a necessary counterpoint to 16S rRNA analysis, which for the Gram-positive phylum has raised as many interesting questions as it has answered.

Cytochromes with spectral characteristics similar to those of the *c*-555 cytochrome from *B. azotoformans* have been reported in low yields from *B. subtilis* [8] and *B. licheniformis* [4,9] grown under conditions of low aeration. Both have a characteristic cytochrome *c*-554 with a split alpha band but so far no sequence information has been available. Cytochromes *c*-551 and *c*-552 from *B. licheniformis* have also been reported [4,9] and the latter has been partially sequenced (Ref. 9; Van Beeumen, personal communication). A complete gene sequence for a *c*-550 cytochrome from *B. subtilis* has also been reported [10]. The only other electron-transfer protein sequence reported from the genus is the *b*-558 cytochrome from *B. subtilis* [2].

Apart from the genus *Bacillus*, *c*-type cytochromes have been reported only from Gram-positive bacteria, to the best of my knowledge, in the aerobic genus *Staphylococcus* [11] which clusters phylogenetically within the *Bacillus* subline in the *Clostridial* (low G + C) subdivision, and in *Mycobacterium phlei* [12] *Microbacterium thermospactum* [13] and *Rhodococcus equi* [14], all of which are aerobic and belong to the *Actionmycetal* (high G + C) subdivision.

Results and Discussion

The *c*-552 and *c*-555 cytochromes from *B. azotoformans* are membrane-bound. They are constitutive and the amount synthesized is approximately the same per unit membrane weight from aerobic and denitrifying cultures. This is in contrast to that reported for the *c*-554 cytochromes from *B. subtilis* and *B. licheniformis*

Correspondence: G.O. Hreggvidsson, Institute of Cell and Molecular Biology, Division of Biological Sciences, University of Edinburgh, Edinburgh, U.K.

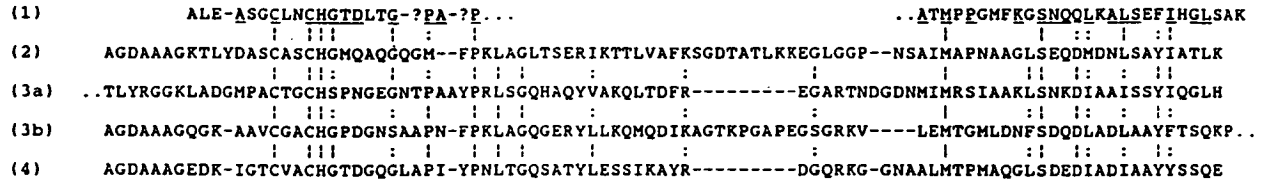


Fig. 1. Alignment of sequence *Bacillus azotoformans* cytochrome *c*-555 with sequences of other bacterial cytochromes *c*. (1) *B. azotoformans* split-alpha cytochrome *c*-555; (2) *Thiobacillus neapolitanus* cytochrome *c*-554(547) [16]; (3a) and (3b) C-terminal and N-terminal halves of *Azotobacter vinelandii* cytochrome c_4 [17]; (4) *Paracoccus halodenitrificans* cytochrome *c*-554(548) [17]. Residues that are identical between sequences are joined by a vertical bar (|) and those that are structurally similar by a colon (:) in what are proposed as signature positions for the class IC sequence class. Residues that are identical in the *B. azotoformans* sequence to any of the other sequences in this figure are underlined.

[3,8,9] which seem to be expressed mainly under conditions of low aeration.

The cytochromes were obtained by tryptic proteolysis from the washed membrane fraction. They were purified by stepwise ammonium sulphate precipitation followed by gel-filtration on Sephadex G-50 and separated by FPLC on a Mono Q column.

Each cytochrome has only two cysteine residues and as deduced from the amino-acid composition the approximate sizes of the cytochromes *c*-552 and *c*-555 are respectively 8 700 and 9 800. This is consistent with these proteins being small functional haem domains cleaved by trypsin treatment from their N-terminal membrane anchors. Such an anchorage is expected to be a likely feature of cytochromes *c* from Gram-positive bacteria, as they do not have periplasmic space and this has recently been confirmed for *c*-550 from *B. subtilis* by gene sequencing [10].

The haem was removed by treatment with $HgCl_2$ in 0.1 M HCl/8 M urea. The apoproteins were desalted and digested with trypsin. The peptides were separated by G-25 gel filtration followed by high-voltage paper electrophoresis. Peptides were sequenced by standard manual and automated techniques [26].

By aligning the partial sequence of *c*-555 from *B. azotoformans* with sequences of well-characterized cytochromes, a marked similarity was found to cytochromes of the IC subclass [15] (Fig. 1). This subclass includes the split-alpha monohaem cytochromes of *Thiobacillus neapolitus* [16] and *Paracoccus halodenitrificans* [17] and both domains of the dihaem cytochrome c_4, which is found in *Azotobacter vinelandii* [17] and various other purple bacteria of the beta and gamma subgroups. The *B. azotoformans* *c*-555 partial sequence shares approx. 25% identity with comparable regions of each of the set of IC beta purple sequences shown in Fig. 1. However, more notable is the retention of the highly conserved pair of prolines following the haem site and the discernible IC signature, with identical or structurally equivalent residues in key positions.

The *c*-552 cytochrome from *B. azotoformans* and the cytochromes *c*-552 and *c*-550 from *B. licheniformis* and *B. subtilis* are clearly homologous (Fig. 2). A significant similarity is between these cytochromes and the ID subclass (*Pseudomonas* *c*-551 cytochromes) [15]; approx. 30% to each of the set of ID sequences shown in Fig. 2. Also noteworthy is the sixth ligand ID signature Met Pro-----Val and the conserved tryptophan close to the C-terminus. In addition, most of the substitutions amongst the consensus of conserved residues of purple bacteria ID cytochromes are conservative. The most notable difference is a deletion of 11 amino acids which seems to correspond to the region at the distal tip of the loop covering the haem crevice [18]. The preceding region (43–54), shows the greatest divergence to the purple bacteria cytochromes. Interestingly, the tryp-

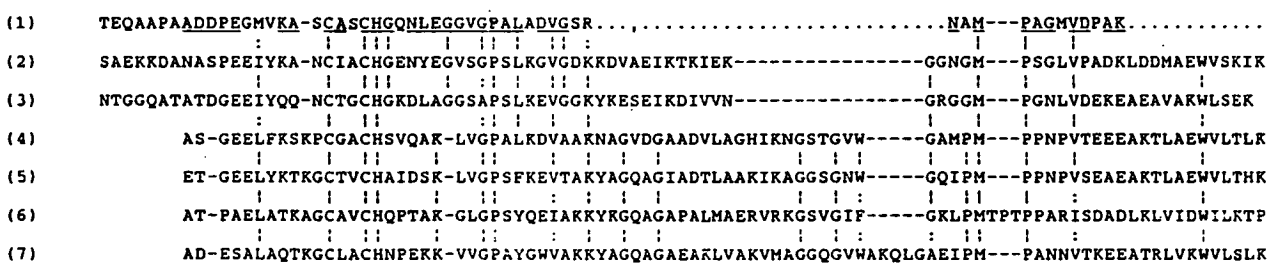


Fig. 2. Alignment of sequence *Bacillus azotoformans* cytochrome *c*-552 with sequences of other bacterial cytochromes *c*. (1) *B. azotoformans* cytochrome *c*-552; (2) *Bacillus subtilis* cytochrome *c*-550 [10]; (3) *Bacillus licheniformans* cytochrome *c*-552 (Ref. 4 and Van Beeumen, J., personal communication; (4) *Pseudomonas mendocina* cytochrome *c*-551 [26]; (5) *Azotobacter vinelandii* cytochrome *c*-551 [27]; (6) *Rhodocyclus tenuis* cytochrome *c*-551 [28]; (7) *Rhodocyclus gelatinosus* cytochrome *c*-551 [28]. Residues that are identical between sequences are joined by vertical bars (|) and those that are structurally similar by a colon (:) in what are proposed as signature positions for the ID sequence. Residues that are identical in the *B. azotoformans* sequence and any of the other sequences in this figure are underlined.

tophan residue −62 is not present. This tryptophan is highly conserved in ID cytochromes and an analogous one is found in different subclasses. The tryptophan is thought to influence the orientation of the haem in the crevice towards the solvent [18] and accords with the observation that the deletion apparently results in the haem being more exposed to the solvent than hitherto encountered in cytochromes *c* of class I.

The occurrence of the IC and ID subclasses in the Gram-positive phylum is unexpected. These subclasses have, so far, been found only in the beta and gamma subgroups of the purple bacteria. Considering the phylogenetic distance between the phyla and the variety of cytochromes *c* of class I, *c*-type cytochromes more specific to the phylum would have been expected.

Oligonucleotide catalogue analysis of 16S rRNA indicates that the ancestral phenotype for gram-positive bacteria was fermentative, similar to the 'true anaerobes' with the deepest branches in the phylum: *Bifidobacteria* of the *Actinomyceatal* division and the clostridia which are found in most sublines of the *Clostridial* subdivision [19]. The occurrence of *c*-type cytochromes in the phylum suggests, however, a much more fully developed respiration mechanism than such an ancestral state would allow for; otherwise, the pathway would be without antecedents in the phylum and would have closer relatives in other phyla. An electron transport pathway of this type and complexity seems either to necessitate a preadapted ancestral system or the acquisition of the cytochromes by lateral gene transfer. In this context it is interesting to note that the species which have been shown to contain cytochromes *c* all belong to phylogenetic clusters which are not particularly deep, which indicates that these groups are not very ancient.

The recent discovery of the Gram-positive photosynthetic species *Heliobacter chlorum* [20], may provide a key to the evolutionary history of the phylum. The available 16S rRNA data for the species [21–23] shows a strong clostridial character and it clusters with *Bacillus* and its relatives [23]. The organism is strictly anaerobic, and cytochromes *c* have been detected on haem-stained SDS gels [20]. On the basis of its discovery, a photosynthetic origin for the Gram-positive bacteria has been proposed [21–23]. However, by the criteria of 16S RNA phylogenetic analysis, it seems that deeper branches must be found to substantiate this claim.

Analysis based on complete sequences of 16S RNA indicates that the Gram-positive bacteria share more recent ancestry with cyanobacteria than with Gram-negative purple bacteria [22]. This does not seem to be supported by the alignment analysis of the *Bacillus* cytochromes above. They have the closest homologues in purple bacteria cytochromes, to the extent of falling into well-defined subclasses from that phylum.

The IC subclass is more closely related to the c_6 subclass (soluble cytochromes *f*) of cyanobacteria than any other, but homology is found only at the C-terminus following the sixth ligand methionine. The c_6 subclass cytochromes show a very high degree of conservation across the great phylogenetic distance between filamentous and unicellular species which have been postulated to have diverged $(1.7–2.0) \cdot 10^9$ years ago [24] and this argues for very ancient sequence characteristics. However, it is possible that the *f* cytochromes may have undergone a major structural change following some fundamental adaption and therefore not retained the N-terminal sequence characteristics seen in the cytochromes from the other two phyla. Therefore the difference between the Gram-positive and the cyanobacterial cytochromes does not necessarily preclude a special phylogenetic relatedness.

Meyer et al. [25] have proposed that functional and structural restraints limit the number of effective substitutions that can occur in a sequence without drastically altering the physiochemical properties of the protein. Thus, with time there should be increasing incidences of convergent mutations until an equilibrium of dissimilarity is reached. Proteins from a range of distantly related species would then remain sequentially equidistant from one another with fundamental characteristics intact. This interpretation would suggest that genes for both the ID and IC cytochromes were present in the ancestor of both the Gram-positive and -negative bacteria and a conservation of functional mode has preserved these specific solutions to the cytochrome fold. However, the alternative explanation, a more recent vertical transmission of the purple bacteria cytochromes by lateral gene transfer into the ancestral species to the genus *Bacillus* and its close relatives, the genus *Staphylococcus* and *H. chlorum,* must also be considered. Such an event may be indicated by the seemingly sporadic occurrence of cytochromes *c* in species in the top branches of the 16S rRNA tree, the genealogical fermentative substructure of the phylum and resultant difficulties of accommodating the photosynthetic species *H. chlorum.* In theory, it should be possible to find evidence for such an event by affinity analysis of the cytochromes from the different purple bacteria subgroups. However, happening at an early stage in the evolutionary history of Gram-positive bacteria, an equilibrium of dissimilarity may still be manifested and obscure specific relationships.

It may seem that, as more sequence information becomes available, the evolutionary history of the Gram-positive bacteria is becoming more elusive and at this stage there is no want of speculative hypotheses. However, complete sequences of 16S rRNA and electron carrier proteins are sparse and with representatives covering the whole metabolic spectrum of Gram-positive bacteria, possibly some, if not all, of the apparent contradictions will be resolved.

Acknowledgements

I thank Richard P. Ambler for valuable comments and encouragement during this work and Linda Kerr and Steven Peacock of the Welmet Protein Characterization Facility and Margret Daniel for skilled and enthusiastic assistance.

References

1 Pichinoty, F., De Barjac, H., Mandel, M., Greenway, B. and Garcia, J.-L. (1976) Ann. Microbiol. 127B, 351–310.
2 Meyer, T.E. and Kamen, M.D. (1982) Adv. Protein Chem. 35, 105–212.
3 Hederstedt, L. (1986) Methods Enzymol. 126, 399–414.
4 Woolley, K.J. (1987) Arch. Biochem. Biophys. 254, 376–379.
5 Miki, K., Sekuzu, I. and Okunuki, K. (1967) Annu. Rep. Biol. Works Fac. Sci. Osaka Univ. 15, 33–58.
6 De Vrij, W., Azzi, A. and Konings, W.N. (1983) Eur. J. Biochem. 131, 97–103.
7 James, W.S., Gibson, F., Taroni, D. and Poole, R.K. (1989) FEMS Microbiol. Lett. 58, 277–282.
8 Miki, K. and Okunuki, K. (1969) J. Biochem. (Tokyo) 66, 831–854.
9 Woolley, K.J. (1984) Ph.D. Thesis, University of Edingburgh.
10 Von Wachenfeldt, C. and Hederstedt, L. (1990) J. Biol. Chem. 265(23), 13939–13948.
11 Keilin, D. (1980) Proc. R. Soc. London. Ser. B. 98, 312.
12 Jacobs, A.J., Kaira, V.K., Cavari, B. and Brodie, A. (1979) Arch. Biochem. Biophys. 194, 531–541.
13 Davidson, C.M. and Hartree, E.F. (1968) Nature (London) 220, 502–505.
14 Reddy, C.A. and Kao, M. (1982) J. Gen. Microbiol. 128, 2379–2383.
15 Ambler, R.P. (1982) in From Cyclotrons to Cytochromes, pp. 263–280, Academic Press, New York.
16 Ambler, R.P., Meyer, T.E., Trudinger, P.A. and Kamen, M.D. (1985) Biochem. J. 227, 1009–1013.
17 Ambler, R.P., Daniel, M., Melis, K. and Stout, C.D. (1984) Biochem. J. 222, 217–227.
18 Salemme, F.R. (1977) Annu. Rev. Biochem. 46, 299–329.
19 Fox, G.E., Stackebrandt, E., Hespell, R.B., Gibson, J., Maniloff, J., Dyer, T.A., Wolfe, R.S., Balch, W.E., Tanner, R., Magrum, L., Zablen, L.B., Gupta, R., Bonen, L., Lewis, B.J., Stahl, D.A., Luehrsen, K.R., Chen, K.N. and Woese, C.R. (1980) Science 209, 457–463.
20 Fuller, R.C., Sprague, S.G., Gest, H. and Blakenship, R.E. (1985) FEBS Lett. 182(2), 345–349.
21 Woese, R.C., Debrunner-Vossbrinck, B., Oyaizu, H., Stackebrandt, E. and Ludwig, W. (1985) Science 229, 762–765.
22 Woese, C.R. (1987) Microbiol. Rev. 51(2), 221–271.
23 Stackebrandt, E., Pohla, H., Kroppenstedt, R., Hippe, H. and Woese, C.R. (1985) Arch. Microbiol. 143, 270–276.
24 Aitken, A. (1988) Methods Enzymol. 167, 145–154.
25 Meyer, T.E., Cusanovich, M.A. and Kamen, M.D. (1986) Proc. Natl. Acad. Sci. USA 83, 217–220.
26 Ambler, R.P. and Wynn, M. (1973) Biochem. J. 131, 485–494.
27 Ambler, R.P. (1973) Syst. Zool. 22, 554–565.
28 Ambler, R.P., Meyer, T.E. and Kamen, M.D. (1979) Nature (London) 278, 661–662.

REFERENCES

Aitken, A. (1976). Protein evolution in cyanobacteria. Nature (London). **26**, pp792-796

Aitken, A. (1988). Protein sequences as taxonomic probes of cyanobacteria. Methods in Enzymology. **167**, pp145-154

Almassy, R.J. and Dickerson, R.E. (1978). 278 Pseudomonas cytochrome c551 at 2.0 A resolution: Enlargement of the cytochrome c family. Proc. Natl. Acad. Sci. USA **75**, pp2674-267

Ambler, R.P. (1963). The amino acid sequence of *Pseudomonas* cytochrome c-551.Biochem. J. **89**, pp349-378

Ambler, R.P. (1971). The amino acid sequence of cytochrome c-551.5 (cytochrome c7). from the green photosynthetic bacterium *Chloropseudomonas ethylica*. FEBS Lett. **18**, pp351-353.

Ambler, R.P. (1976). Standards and accuracy in amino acid sequence determination. In: 'Structure-Function Relationships of Proteins' (Markham, R., Horne, R.W. eds.). Elsvier-North Holland, Amsterdam. pp1-14

Ambler, R.P. (1977a). Amino acid sequences and bacterial phylogeny. In: ´Evolution of Protein Molecules´ (Matsubara,H., Yamanaka, T. eds.). Japan Scientific Societies Press. pp311-322

Ambler, R.P. (1977b). Cytochrome c and copper protein evolution in prokaryotes. In: ´The evolution of metalloenzymes, metalloproteins and related materials´ (Leigh, G.J., ed.). Symposium Press, London. pp100-118

Ambler, R.P. (1982). The structure and classification of cytochromes c. In; ´From Cytochromes to Cyklotrons´(N.O. Kaplan and A. Robinson, eds.). Academic Press, London. pp263-279

Ambler, R.P. (1985). Protein sequencing and taxonomy. In: ´Twenty Five Years of Numerical Taxonomy ´ (Jones, D., Goodfellow, M. and Priest, F.G., eds.). Academic Press, London. pp307-335

Ambler, R.P. (1991). Sequence variability in bacterial cytochromes c. Biochimica et Biophysica Acta **1058**, pp42-4

Ambler, R.P. and Bartsch, R.G. (1975). Amino acid sequence similarity between cytochrome f from a blue green bacterium and from algal chloroplasts. Nature (London). **253**, pp285-299

Ambler, R.P. and Wynn, M. (1973). The amino acid sequences of cytochromes c-551 from three species of *Pseudomonas*. Biochem. J. **131**, pp484-498

Ambler R.P., Daniel M., Hermoso J., Meyer T.E., Bartsch R.G. and Kamen M.D. (1979a). Cytochrome c2 sequence variation among the recognised species of purple nonsulphur photosynthetic bacteria. Nature, **278**, pp659-660

Ambler, R.P., Meyer, T.E. and Kamen, M.D. (1979b). Anomalies in amino acid sequences of small cytochromes c and cytochromes c' from two species of purple photosynthetic bacteria. Nature (London). **278**, pp661-662

Ambler, R.P., Daniel, M., Melis, K. and Stout, C.D. (1984). The amino acid sequence of the dihaem cytochrome c4 from the bacterium *Azotobacter vinelandii*. Biochem. J. **222**, pp217-222

Ambler, R.P., Meyer, T.E., Trudinger, P.A. and Kamen, M.D. (1985). The amino acid sequence of the cytochrome c-554(547). from the chemolithotropic bacterium *Thiobacillus neapolitanus*. Biochem. J. **227**, pp1009-1013

Ambler, R.P., Dalton, H., Meyer, T.E., Bartsch, R.G. and Kamen, M.D. (1986). The amino acid sequence of cytochrome c-555 from the methane oxidising bacterium *Methylococcus capsulatus*. Biochem. J. **233**, pp333-337

Anfinsen, C.B. (1959). 'The Molecular Basis of Evolution.' John Wiley & Sons, New York.

Anraku, Y. (1987). Bacterial electron transport chains. Ann. Rev. Biochem. **57**, pp101-132

Bartsch, R.G. (1991). The distribution of soluble metallo-redox proteins in purple phototrophic bacteria. Biochemica et Biophysica Acta **1058**, pp28-30

Bartsch, R.G., Ambler, R.P., Meyer, T.E. and Cusanovich, M.A. (1989). Effect of aerobic growth conditions on the soluble cytochrome content of the purple phototrophic bacterium *Rhodobacter sphaeroides*: Induction of cytochrome c554. Archives of Biochemistry and Biophysics **271**, pp433-440

Bhatia, G.E., Finzel, B.C. and Kraut, J. (1984). 2A resolution X-ray coordinates of *R. rubrum* cytochrome c2. Deposited with the protein data bank.

Blundell, T., Carney, D., Gardner, S., Hayes, F., Howlin, B., Hubbard, T., Overington, J., Singh, D.A., Sibanda, B.L. and Sutcliffe, M. (1988). Knowledge-based protein modelling and design. Eur. J. Biochem. **172**, pp513-520

Carter, D.C., Melis, K.A., O'Donnell, S.E., Burgess, B.K., Furey, W.F., Wang, B.C. and Stout, C.D. (1985). Crystal structure of *Azotobacter* cytochrome c5 at 2.5 A resolution. J. Mol. Biol. **184**, pp279-295

Chothia, C. and Lesk, A.M. (1984). Helix movements and the reconstruction of the haem pocket during the evolution of the cytochrome c family. J. Mol. Biol. **182**, pp151-158

Chothia, C. and Lesk, A.M. (1986). The relation between the divergence of sequence and structure in proteins. The EMBO Journal **5**, pp823-826

Chothia, C. and Lesk, A.M. (1987). The Evolution of Protein Structure. Cold Spring Harbor Symposia on Quantitative Biology, LII, pp399-405

Claus D. and Berkeley, R.C.W. (1986). Genus Bacillus Cohn 1872, 174. In: Bergey's Manual of Systematic Bacteriology' (Williams and Wilkins, Baltimore). vol2, pp1105-1139

Clinton Fuller, R., Sprague, S.G., Gest, H. and Blankenship, R.E. (1985). A unique photosynthetic reaction center from *Heliobacterium chlorum.* FEBS 2361 **182**, pp345-349

Crick, F.H.C. (1958). On protein synthesis. Symposium of the Society for Experimental Biology **12**, pp138-163

Daldal, F., Cheng, S., Applebaum, J., Davidson, E. and Prince, R.C. (1986). Cytochrome c2 is not essential for photosynthetic growth of *Rhodopseudomonas capsulata.* Proc. Natl. Acad. Sci. USA **83**, pp2012-2016

Dayhoff, M.O., Scwartz, R.M. and Orcutt, B.C. (1978). A model of evolutionary change in proteins. In: 'Atlas of Protein Sequence and Structure' National Biomedical Research Foundation, Washington, D.C. pp345-352

Dayhoff, M.O., Barker, W.C. and Hunt, L.T. (1983). Establishing homologies in protein sequences. Methods in enzymology **91**, pp524

Denariaz, G., Payne, W.J. and LeGall, J. (1989). A halophilic denitrifier, *Bacillus halodenitrificans* sp. nov. International Journal of Systematic Bacteriology **39**, pp145-151

Denariaz, G., Payne, W.J. and LeGall, J. (1991). The denitrifying nitrite reductase of *Bacillus halodenitrificans.* Biochimica et Biophysica Acta 1056, pp225-232

DeVries, W., van Wijck-Kapteyn, M.C. and Oosterhuis, S.K.H. (1974). The presence and function of cytochromes in *Selenomonas ruminantium, Anaerovibrio lipolytica and Veillonella alcalescens.*Journal of General Microbiology **81**, pp69-78

De Vrij, W., Azzi,A. and Konings, W.N. (1983). Structural and functional properties of cytochrome c oxidase from *Bacillus subtilis* W23. Eur. J. Biochem. **131**, pp97-103

De Wachter,R., Huysmans, E. and Vandenberg, H.E. (1985). 5S ribosomal RNA as a tool for studying evolution. In: 'Evolution of Prokaryotes' (Schleifer, K.H., Stackebrandt, E., eds.). Academic Press, New York. pp115-141

Dickerson, R.E. (1971). The structure of cytochrome c and the rates of molecular evolution. J. Molec. Evol. **1**, pp26-45

Dickerson, R.E. (1980). Cytochrome c and the evolution of energy metabolism. Sci. Amer. **242**, pp136-153

Eisenberg, D., Wesson, M. and Wilcox, W.(1989). Hydrophobic moments as tools for analyzing protein sequences and structures. In:' Prediction of Protein Structure And The Principle of Protein Conformation' (Fasman,G.D. ed.). Plenum Press, New York. pp635-646

Fasman, G.D. (1989). The development of the prediction of protein structure. In: ´Prediction of Protein Structure And The Principle of Protein Conformation´ (Fasman, G.D. ed.). Plenum Press, New York. pp193-316

Fee, J.A., Mather, M.W., Springer, P., Hensel, S. and Buse, G. (1988). Isolation and partial sequence of the A-protein gene of *Thermus thermophilus* cytochrome c1aa3. Annals New York Academy of Sciences, **550**, pp33-38

Felsenstein, J. (1988). Phylogenies from molecular sequences: Inference and reliability. Ann. Rev. Genet. **22**, pp521-565

Findley, J.B. (1987). Isolation and labelling of membrane protein and peptides. In: ´Biological Membranes´ (FindleyJ.B., Evans, W.H., eds.). IRL Press, Oxford.

Fitch, W.M. and Margoliash, E. (1967). Construction of phylogenetic tree. Science **155**, pp279-284

Ford, P., personal communication.

Fox, G.E., Stackebrandt, E., Hespell, R.B., Gibson, J., Maniloff, J., Dyer, T.A., Wolfe, R.S., Balch, W.E., Tanner, R.S., Magrum, L.J., Zablen, L.B., Blakemore, R., Gupta, R., Bonen, L., Lewis, B.J., Stahl, D.A., Luehrsen, K.R., Chen, K.N. and Woese, C.R. (1980). The phylogeny of prokaryotes. Science **209**, pp457-463

Fujiwara ,Y., Oka, M., Hamamoto, T. and Sone, N. (1993). Cytochrome c-551 of the thermophilic bacterium PS3, DNA sequence and the mature cytochrome. Biochimica et Biophysica Acta. in press

Gest, H., personal communication.

Garnier, J. and Robson, B. (1989). The GOR method for prediction secondary structure in proteins. In: ´Prediction of Protein Structure And The Principles of Protein Conformation´ (G.D:Fasman, ed.). Plenum Press, New York. pp417-466

Gillespie, J.H. (1986). Natural selection and the molecular clock. Mol. Biol. Evol. **3**, pp138-155

Goodhew, C.F., Brown, K.R. and Pettigrew, G.W. (1986). Haem staining in gels, a useful tool in the study of bacterial c-type cytochromes. Biochimica et Biophysica Acta **852**, pp288-294

Gordon, R.E. (1981). One hundred and seven years of the genus *Bacillus* . In: ´The Aerobic Endospore-forming Bacteria: Classification and Identification´ (Berkeley, R.C., Goodfellow, WM eds.). Academic Press, London. pp1-16

Gribskov, M., McLachlan, A.D. and Eisenberg, D. (1987). Profile analysis: Detection of distantly related proteins. Proc. Natl. Acad. Sci. USA **84**, pp4355-4358

Gribskov, M., Luthy, R. and Eisenberg, D. (1990). Profile analysis. Methods in enzymology **183**, pp146-158

Hames, D. (1990). One-dimensional polyacrylamide gel electrophoresis. In: ´Gel Electrophoresis of Proteins´ (Hames, B.D., and Rickwood, D.,eds.). IRL Press, Oxford. pp1-149

Hampsey, D.M., Dast, G. and Sherman, F. (1986). Amino acid replacements in yeast iso-1-cytochrome c. The Journal of Biological Chemistry **261**, pp3259-3271

Hon-Nami, K. and Oshima,T. (1978). C-type cytochromes isolated from an extreme thermophile, *Thermus thermophilus* HB8. J. Biochem. **83**, pp629-631

Hon-nami, K. and Oshima, T. (1977). J. Biochem. **82**, pp769-776

Hori, H. and Osawa, O. (1986). Evolutionary change in 5SrRNA secondary structure and a phylogenic tree of 352 5SrRNA species. BioSystems. **19**, pp163-172.

Ishizuka, M., Machida, K., Shimada, S., Mogi, A., Tsuchiya, T., Ohmori, T., Souma, Y., Gonda, M. and Sone, N. (1990). Nucleotide sequence of the gene coding for four subunits of cytochrome c oxidase from the thermophilic bacterium PS3. J. Biochem. **108**, pp866-873

Jones, C.W. (1985). The evolution of bacterial respiration. In: ´Evolution of Prokaryotes´ (Schleifer, K.H. and Stackebrandt, E. eds.). Academic Press, London. pp175-204

Jones, D. and Krieg, N.R. (1986). Serology and chemotaxonomy. In: ´Bergey´s Manual of Systmatic Bacteriology´ (Sneath, P.H.A. ed.). Williams and Wilkins, Baltimore. vol2, pp979-987

Jones, D. and Sneath, P.H.A. (1970). Genetic transfer and bacterial taxonomy. Bacteriological Reviews **34**, pp40-81

Kallas, T., Spiller, S. and Malkin, R. (1988). Primary structure of cotranscribed genes encoding the Rieske Fe-S and cytochrome f proteins of the cyanobacterium Nostoc PCC 7906. Proc. Natl. Acad. Sci. USA **85**, pp5794-5798

Kamen M.D., Errede, B.J. and Meyer, T.E. (1977). In: 'Evolution of Protein Molecules´. (H. Matsubara, T. Yamanaka, eds.). Japan Scientific Societies Press', Tokyo.

Keilin, D. (1925). On cytochrome, a respiratory pigment, common to animals, yeast and higher plants. Proc. Roy. Soc. London B **98**, pp312-339

Keilin, D (1970). ´The History of Cell Respiration and Cytochrome c´ (Keilin, J. prep.). Cambridge University Press, Cambridge.

Kimura, M. (1983). ´The neutral theory of molecular evolution´. Cambridge University Press, Cambridge.

Korszun, Z.R. and Salemme, F.R. (1977). Structure of cytochrome c555 of *Chlorobium thiosulfatophilum:* Primitive low-potential cytochrome c. Proc. Natl. Acad. Sci. USA **74**, pp5244-5247

Krawiec, S. and Riley, M. (1990). Organization of the bacterial chromosome. Microbiological Reviews **54**, pp502-539

Kutoh, E. and Sone, N. (1988). Quinol-cytochrome c oxidoreductase from the thermophilic bacterium PS3. The Journal of Biological Chemistry **263**, pp9020-9026

Lai, J-S., Sarvas, M., Brammar, W.J., Neugebauer, K. and Wu, H.C.(1981). *Bacillus licheniformis* penicillinase synthesized in *Escherichia coli* contains covalently linked fatty acid and glyceride. Proc. Natl. Acad. Dci. USA **78**, pp3506-3510

Lake, J.A. (1988). Origins of the eukaryotic nucleus determined by rate invariant analysis of rRNA sequences. Nature (London). **331**, pp184-186

Lake, J.A. (1991). Tracing origins with molecular sequences: Metazoan and eukaryotic beginnings. TIBS **16**, pp46-50

Lanyi, J.K. (1968). Studies of the electron transport chain of extremely halophilic bacteria. Archives of Biochemistry and Biophysics **128**, pp716-724

Lathe, L. (1985). Synthetic oligonucleotide probes deduced from amino acid sequence data. J. Mol. Biol. **183**, pp1-12

Lauraeus, M., Haltia, T., Saraste, M. and Wikström, M. (1991). Eur. J. Biochem. **197**, pp699-705

Laemmli, U.K., (1970). Cleveage of structural proteins during the assembly of head of bacteriophage T4. Nature **277**, pp680-685.

Linerud, K., personal communication.

Logan, N.A. and Berkeley, R.C.W. (1981). Classification and identification of members of the genus *Bacillus* using API tests. In: ´The Aerobic Endospore-forming Bacteria: Classification and Identification´ (Berkeley, R.C.W. and Goodfellow, M., eds). Academic Press, London. pp105-140

Loprasert S., Negoro, S. and Okada, H. (1989). Cloning, nucleotide sequence and expression in *Escherechia coli* of the *Bacillus stearothermophilus* peroxidase gene (*per*A). Journal of Bacteriology 171, pp4871-4875

Ludwig, M.L., Pattridge, K.A., Powers, T.B., Dickerson, R.E. and Takano, T. (1982). Structure analysis of a ferricytochrome c from the cyanobacterium *Anacystis nidulans.* In: ´Electron Transport and Oxygen Utilization´ (Ho, C., ed.). Elsevier, Amsterdam. pp27-32

Margoliash, E., Ferguson-Miller, S., Brautigan, D.I. and Chaviano, A.H. (1976a). Functional basis for evolutionary change in cytochrome c structure. In: ´Structure-Function Relationships of Proteins´ (Markham, R., Horne, R.W., eds.). Elsevier-North Holland, Amsterdam. pp145-165

Margoliash, E., Ferguson-Miller, S., Kang, C.H. and Brautigan, D.L. (1976b). Do evolutionary changes in cytochrome c structure reflect structural adaptations? Fed. Proc. **35**, pp2124-2130

Mather, M.W., Springer, P. and Fee, J.A. (1991). Cytochrome oxidase genes from *Thermus thermophilus.* The Journal of Biological Chemistry **266**, pp5025-5035

Matsuura, Y., Takano, T. and Dickerson, R.E. (1982). Structure of cytochrome c-551 from *Ps. aeroginosa* refined at 1.6A resolution and comparison of the two redox forms. J. Mol. Biol. **156**, pp389-409

Maynard-Smith, J., Dowson, C.G. and Spratt, B.G. (1991). Localized sex in bacteria. Nature **349**, pp29-31

McLachlan, A.D. (1971). Tests for comparing related amino-acid sequences. cytochrome c and cytochrome c551. J. Mol. Biol. **61**, pp409-424

McLachlan, A.D.(1987). Gene duplication and the origin of repetitive protein structures. Cold Spring Symposia on Quantitative Biology, LII, pp411-420

Meyer, T.E. (1980). Cytochromes and ferredoxins in bacterial electron transport and classification. In: 'From Cyclotrons to Cytochromes' (Kaplan, N.O., Robinson, A.B. eds.). Academic Press, New York. pp157-180

Meyer, T.E. (1991). Evolution of cytochromes and photosynthesis. Biochemica et Biophysica Acta **1058**, pp31-34

Meyer, T.E. and Kamen, M.D. (1982). New perspectives on c-type cytochromes. Adv. Prot. Chem. **35**, pp105-212

Meyer, T.E., Cusanovich, M.A. and Kamen, M.D. (1986). Evidence against use of bacterial amino acid sequence data for construction of all-inclusive phylogenetic trees. Proc. Natl. Acad. Sci. USA **83**, pp217-220

Miki, K. and Okunuki, K. (1969a). Cytochromes of *Bacillus subtilis;* Purification and spectral properties of cytochrome c550 and c554. The Journal of Biochemistry **66**, pp831-845

Miki, K. and Okunuki, K. (1969b). Cytochromes of *Bacillus subtilis*; Physicochemical and enzymatic properties of cytochrome c550 and c554. The Journal of Biochemistry **66**, pp845-854

Minnikin D.E. and Goodfellow, M. (1981). Lipids in the classification of *Bacillus* and related Taxa. In: 'The Aerobic Endospore-forming Bacteria: Classification and Identification' (Berkeley, R.C.W.,Goodfellow, M., eds.). Academic Press, London. pp59-90

Moazed, D. and Noller, H.F. (1987). Interaction of antibiotics with functional sites in 16S ribosomal RNA. Nature **327**, pp389-394

Moore, G.R. and Pettigrew, G.W. (1990). Cytochromes C; Evolutionary, Structural and Physicochemical Aspects (Williams, R.J.P. ed.). Springer-Verlag, London.

Murrey, R.G.E. (1986). The higher taxa, or a place for everything? In: 'Bergey's Manual of Systematic Bacteriology' (Sneath, P.H.A. ed.). Williams and Wilkins, Baltimore. vol2, pp995-998

Nakagawa, A., Higuchi, Y., Yasuoka, N., Katsube, Y. and Yagi, T. (1990). S-class cytochromes c have a variety of folding patterns: Structure of cytochrome c-553 from *Desulfovibrio vulgaris* determined by the multi-wavelength anomalous dispersion method. J. Biochem. **108**, pp701-703

Nei, M. (1987). 'Molecular Evolutionary Genetics'. Columbia University Press, New York.

Nielsen, J.B.K., Caulfield, M.P. and Lampen, J.O. (1981). Lipoprotein nature of *Bacillus licheniformis* membrane penicillinase. Proc. Natl. Acad. Sci. USA **78**, pp3511-3515

Nordling, M., Young, S., Karlsson, B.G. and Lundberg, L.G. (1990). The structural gene for cytochrome c551 from *Pseudomonas aeruginosa*. FEBS Letters **259**, pp230-232

Norris, J.R. (1981). *Sporosarcina* and *Sporolactobacillus*. In: 'The Aerobic Endospore-forming Bacteria: Classification and Identification' (Berkeley, R.C.W. and Goodfellow, M., eds.). Academic Press, London. pp337-358

O'Donnell, A.G. and Norris, J.R. (1981). Pyrolysis gas-liquid chromatographic studies in the genus *Bacillus*. In: 'The Aerobic Endospore-forming Bacteria: Classification and identification' (Berkeley, R.C.W., Goodfellow, M., eds.), Academic Press, London. pp141-180

Ochman, H. and Wilson, A.C. (1987). Evolution in bacteria: Evidence for a universal substitution rate in cellular genomes. J. Mol. Evol. **26**, pp74-86

Pace, C.N. (1975). The stability of globular proteins. Crit. Rev. Biochem. **3**, pp1-43

Pace, N.R., Stahl, D.A., Lane, D.J. and Olsen, G.J. (1985). ASM News **51**, pp4-12

Patthy, L. (1987). Detecting homology of distantly related proteins with consensus sequences. J. Mol. Biol. **198**, pp567-577

Pettigrew, G.W. (1974). The purification and amino acid sequences of cytochrome c-552 from *Euglena gracilis*. Biochem. J. **139**, pp449-459

Pettigrew, G.W. (1991). The cytochrome c peroxidase of *Paracoccus denitrificans*. Biochimica et Biophysica Acta **1058**, pp25-27

Pettigrew, G.W. and Brown, K.B. (1988). Free and membrane-bound forms of bacterial cytochrome c. Biochem. J. **252**, pp427-435

Pettigrew, G.W. and Meyer, T.E. (1971). Biochem. J. **125**, pp46

Pettigrew, G.W. and Moore, G.E. (1987). 'Cytochromes c: Biological Aspects'. Springer-Verlag, New York Heidelberg.

Pichinoty, P., de Barjac, H., Mandel, M., Greenway, B. et Garcia, J.-L. (1976). Une nouvelle bacterie sporulee, denitrifiante, mesophile: *Bacillus azotoformans* n.sp. Ann. Microbiol. (Inst. Pasteur). **127 B**, pp351-361

Pichinoty, F., Garcia, J.-L., Job, C. et Durand, M. (1978). La denitrification chez *Bacillus licheniformis*. Can. J. Microbiol. **24**, pp45-49

Pichinoty, F., Mandel, M. and Garcia, J.-L (1979). The properties of novel mesophilic denitrifying bacillus cultures found in tropical soils. Journal of General Microbiology **115**, pp419-430

Priest, F.G., Goodfellow, M. and Todd, C. (1981). The genus *Bacillus*: a numerical analysis. In: 'The Aerobic Endospore-forming Bacteria:

Classification and Identification' (R.C.W. Berkeley, M. Goodfellow, eds.). Academic Press, London. pp91-104

Qureshi, M.H., Yumoto, I., Fujiwara, T., Fukomori, Y. and Yamanaka, T. (1990). A novel aco-type cytochrome-c oxidase from a facultative alkalophilic bacillus:
Purification and some molecular and enzymatic features. J. Biochem. 107, pp480-485

Richardson, J.S. and Richardson, D.C. (1989). Principles and patterns of protein conformation. In: 'Prediction of protein structure and the principle of protein conformation' pp.1-98 (G.D. Fasman, ed.). Plenum Press, New York. pp1-98

Roe, B.A., Ma, D.-P., Wilson, R.K. and Wong, J.F.-H. (1985). J. Biol. Chem. 260, pp9759-9774

Salemme, F.R., Kraut, J. and Kamen, M.D. (1973). Structural bases for function in cytochromes c. The Journal of Biological Chemistry **248**, pp7701-7716

Salemme, F.R. (1977). Structure and function of cytochromes c. Ann. Rev. Biochem. 4, pp299-329

Sali, A., Overington, J.P., Johnson, M.S. and Blundell, T.L. (1990). From comparisons of protein sequences and structures to protein modelling and design. TIBS 15, pp235-240

Sanbongi, Y., Ishii, M., Igarashi, Y. and Kodama, T. (1989). Amino acid sequence of cytochrome c-552 from a thermophilic hydrogen-oxidizing bacterium, *Hydrogenobacter thermophilus.* Journal of Bacteriology **171**, pp65-69

Sanbongi, Y., Yang, J.-H., Igarashi, Y. and Kodama, T. (1991). Cloning nucleotide sequence and expression of the cytochrome c-552 gene from *Hydrogenobacter thermophilus.* Eur. J. Biochem. **198**, pp7-12

Santana, M., Kunst, F., Hullo, M.F., Rapoport, G., Danchin, A. and Glaser, P. (1992). Molecular cloning, sequencing and physiological characterization of the qox operon from *Bacillus subtilis* encoding the aa3-600 quinol oxidase. The Journal of Biological Chemistry **267**, pp10225-10231

Saraiva, L.M., Denariaz, G., Liu, M.Y., Payne, W.J., LeGall, J. and Moura, I. (1992). NMR and EPR studies on a monoheme cytochrome c550 isolated from *Bacillus halodenitrificans.* Eur. J. Biochem. **204**, pp1131-1139

Schleifer, K.H. and Stackebrandt, E. (1983). Molecular systematics of prokaryotes. Ann. Rev. Microbiol. **37**, pp143-187

Scherer, S. (1990). Do photosynthetic and respiratory electron transport chains share redox proteins?. Trends in Biochem. Sci. **15**, pp458-462

Schulp, J.A. and Stouthamer, A.H. (1970). The influence of oxygen, glucose and nitrate upon the formation of nitrate reductase and the respiratory system in *Bacillus licheniformis.* Journal of General Microbiology **64**, pp195-203

Schwartz, R.M. and Dayhoff, M.O. (1978a). Cytochromes. In: 'Atlas of Protein Sequence and Structure' (Dayhoff, M.O., ed.). National Biomedical Research Foundation, Washington, D.C. vol5, suppl.3, pp29-52

Scwartz, R.M. and Dayhoff M.O. (1978b). Matrices for detecting distant relationships. In: 'Atlas of Protein Sequence and Structure' (Dayhoff, M.O., ed.). National Biomedical Research Foundation, Washington, D.C. pp353-358

Sidow, A. and Wilson, A.C. (1990). Compositional statistics: An improvement of evolutionary parsimony and its application to deep branches in the tree of life. J. Mol. Evol. **31**, pp51-68

Smith, G.B. and Tiedje, J.M. (1992). Isolation and characterization of a nitrite reductase gene and its use as a probe for denitrifying bacteria. Applied and Environmental Microbiology **58**, pp376-384

Sneath, P.H.A. (1989). Analysis and interpretation of sequence data for bacterial systematics: the view of a numerical taxonomist. System. Appl. Microbiol: **12**, pp15-31

Sneath, P.H.A. and Sokal, R.R. (1973). 'Numerical Taxonomy'. Freeman, San Francisco.

Sone, N. and Fujiwara, Y. (1991). Haem O can replace haem A in the active site of cytochrome c oxidase from thermophilic bacterium PS3. FEBS 010069 **288**, pp154-158

Sone, N. and Yanagita, Y. (1982). A cytochrome aa3-type terminal oxidase of a thermophilic bacterium purification, properties and proton pumping. Biochimica et Biophysica Acta **682**, pp216-226

Sone, N., Kagawa, Y. and Orii, Y. (1983). J. Biochem. **93**, pp1329-1336

Sone, N., Sekimachi, M. and Kutoh, E. (1987). J. Biol. Chem. **262**, pp15386-15391

Sone, N., Yanagita, Y., Hon-nami, K., Fukumori, Y. and Yamanaka, T. (1983). Proton-pumping activity of *Nitrobacteria gilis* and *Thermus thermophilus* cytochrome c oxidases. FEBS Lett. **155**, pp150-154

Sone, N., Sekimachi, M. and Kutoh, E. (1987). Identification and properties of a quinol oxidase super-complex composed of a bc1 complex and cytochrome oxidase in the thermophilic bacterium PS3. The Journal of Biological Chemistry **262**, pp15386-15391

Sone, N., Yokoi, F., Fu, T., Ohta, S., Metso, T., Raitio, M. and Saraste, M. (1988). Nucleotide sequence of the gene coding for cytochrome oxidase subunit I from the thermophilic bacterium PS3. J. Biochem. **103**, pp606-610

Sone, N., Kutoh, E. and Yanagita, Y. (1989). Cytochrome c-551 from the thermophilic bacterium PS3 grown under air-limited conditions. Biochimica et Biophysica Acta **977**, pp329-334

Sone, N., Kutoh, E. and Sato, K. (1990). A cytochrome O-type oxidase of the thermophilic bacterium PS3 grown under air-limited conditions. J. Biochem. **107**, pp597-502

Sonea, S. (1989). A new look at bacteria. ASM News **55**, pp584-585

Stackebrandt, E. (1985). Phylogeny and phylogenetic classification of prokaryotes. In:´ Evolution of Prokaryotes´ (Schleifer, K.H., Stackebrandt, E., eds.). Academic Press, London. pp309-334

Stackebrandt, E. and Woese C.R. (1981). The evolution of prokaryotes. In: ´Molecular And Cellular Aspects of Microbial Evolution´ (Carlile, M.J., Collins, J.F., Moseley, B.E.B. eds.). Cambridge University Press, Cambridge. pp1-31

Stackebrandt, E, Pohla, H, Kroppenstedt, R, Hippe, H. and Woese, C.R (1985). 16S rRNA analysis of Sporomusa, Selenomonas, and Megasphaera: On the phylogenetic origin of Gram-positive eubacteria. Arch Microbiol **143**, pp270-276

Stouthamer, A.H. (1992). Metabolic pathways in *Paracoccus dinitrificans* and closeley related bacteria in relation to the phylogeny of prokaryotes. Antonie van Leeuwenhoek Microbiol. Serol., **61**, pp1-14

Swofford, D.L and Olsen, G.J. (1990). Phylogeny reconstruction. In: ´Molecular Systematics´ (Hillis, D.M., and Moritz, C., eds.). Sinauer Associates, Inc., Sunderland, Mass. pp411-501

Taylor, W.R. (1986). Consensus template alignment. J. Mol. Biol. **188**, pp233-258

Taylor, W.R. (1990). Hierarchial method to align large numbers of biological sequences. Methods in Enzymology **183**, pp456-469

Thatcher, D. and Ambler R.P., personal communication

Thauer, R.K., Jungermann, K. and Decker, K. (1977). Energy conservation in chemotrophic anaerobic bacteria. Bacteriological Review **41**, pp100-180

Titani, K., Ericsson, L.H., Hon-nami, K. and Miyazawa, T. (1985). Amino acid sequence of cytochrome c-552 from *Thermus thermophilus* HB8. Biochem. Biophys. Res. Commun. **128**, pp781-787

Towe, K.M. (1990). Aerobic respiration in the Archaean? Nature **348**, pp54-57

Trumpower, B.L. (1990). Cytochrome bc_1 Complexes of Microorganisms. Microbiological Reviews **54**, pp101-129

van Beeumen, J., personal communication.

van Beeumen, J. (1991). Primary structure diversity of prokaryotic diheme cytochromes c. Biochimica et Biophysica Acta **1058**, pp56-60

van Beeumen, J., Van Bun, S., Meyer, T.E., Bartsch, R.G. and Cusanovich, M.A. (1990). Complete amino acid sequence of the cytochrome subunit and amino-terminal sequence of the Flavin subunit of Flavocytochrome c (Sulfide Dehydrogenase). from *Chlorobium thiosulfatophilum.* The Journal of Biological Chemistry **265**, pp9793-9799

van 'T Riet, J., Wientjes, F.B., Van Doorn, J. and Planta, R.J. (1979). Purification and characterization of the respiratory nitrate reductase of *Bacillus licheniformis.*Biochimica et Biophysica Acta **576**, pp347-360

von Wachenfeldt, C. and Hederstedt, L. (1990). *Bacillus subtilis* 13-Kilodalton cytochrome c-550 encoded by cccA consists of a membrane-anchor and a heme domain. The Journal of Biological Chemistry **265**, pp13939-13948

Walker, J.E., Saraste, M. and Gay, N.J. (1984). The UNC operon-nucleotide sequence regulation and structure of ATP synthase. Biochimica et Biophysica Acta. **768**, pp164-200

Weisburg, W.G., Giovannoni, S.J. and Woese, C.R. (1989). The *Deinococcus-thermus phylum* and the effect of rRNA composition on phylogenetic tree construction. System. Appl. Microbiol. **11**, pp128-134

Williams, R.A.D. (1992). The genus Thermus. In: 'Thermophilic bacteria' (Kristjánsson, J.K., ed.). CRC Press, Inc., Florida. pp51-62

Wilson, A.C, Ochman, H. and Prager, E.M (1987). Molecular time scale for evolution. Trends in Genetics **3**, pp241-247

Woese, C.R (1985). Why study evolutionary relationships among bacteria? In:' Evolution of Procaryotes' (Schleifer, K.H, Stackebrandt, E. eds.). Academic Press, London. pp1-30

Woese, C.R. (1987). Bacterial evolution. Microbiol. Rev. **51**, pp221-271

Woese, C.R., Debrunner-Vossbrinck, B.A., Oyaizu, H., Stackebrandt, E. and Ludwig, W. (1985). Gram-positive bacteria: possible photosynthetic ancestry. Science **229**, pp762-765

Wood, P.M. (1983). Why do c-type cytochromes exist? FEBS Lett. **164**, pp223-226

Wood, P.M. (1991). Why do c-type cytochromes exist? - Reprise. Biochimica et Biophysica Acta **1058**, pp5-7

Wood, W.I., Gitschier, J., Lasky, L.A. and Lawn, R.M. (1985). Base composition-independent hybridization in tetramethylammonium chloride: A method for oligonucleotide screening of highly complex gene libraries. Proc. Natl. Acad. Sci. USA **82**, pp1585-1588

Woolley, K.J. (1984). Ph.D. Thesis, University of Edinburgh.

Woolley, K.J. (1987). The c-type cytochromes of the Gram-positive bacterium *Bacillus licheniformis*. Archives of Biochemistry and Biophysics **254**, pp376-379

Yaoi, H. and Tamiya, H. (1928). On the respiratory pigment, cytochrome, in bacteria. Proc. Imp. Acad. Japan **4**, pp436-440

Zhao, H., Yang, D., Woese, C.R. and Bryant, M.P. (1990). Assignment of *Clostridium bryantii* to *Syntrophospora bryantii* gen. nov., comb. nov. on the basis of a 16S rRNA sequence analysis of its crotonate-grown pure culture. International Journal of Systematic Bacteriology, **40**, pp40-44

Zuckerkandl, E. (1976). Evolutionary processes and evolutionary noise at the molecular level. J. Mol. Evol. **7**, pp269-312

Zuckerkandl, E. and Pauling, L. (1965). Molecules as documents of evolutionary history. J. Theor. Biol. **8**, pp357-366